

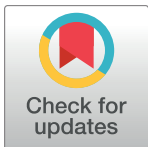
RESEARCH ARTICLE

Can the Immune System Perform a t-Test?

Bruno Filipe Faria^{1,2,3}, Patricia Mostardinha^{1,3}, Fernao Vistulo de Abreu^{1,2,3*}

1 Department of Physics, University of Aveiro, Aveiro, Portugal, **2** Institute for Biomedicine - iBiMED, University of Aveiro, Aveiro, Portugal, **3** I3N Institute for Nanostructures, Nanomodelling and Nanofabrication, Aveiro, Portugal

* fva@ua.pt



Abstract

The self-nonsel self discrimination hypothesis remains a landmark concept in immunology. It proposes that tolerance breaks down in the presence of nonself antigens. In strike contrast, in statistics, occurrence of nonself elements in a sample (i.e., outliers) is not obligatory to violate the null hypothesis. Very often, what is crucial is the combination of (self) elements in a sample. The two views on how to detect a change seem challengingly different and it could seem difficult to conceive how immunological cellular interactions could trigger responses with a precision comparable to some statistical tests. Here it is shown that frustrated cellular interactions reconcile the two views within a plausible immunological setting. It is proposed that the adaptive immune system can be promptly activated either when non-self ligands are detected or self-ligands occur in abnormal combinations. In particular we show that cellular populations behaving in this way could perform location statistical tests, with performances comparable to t or KS tests, or even more general data mining tests such as support vector machines or random forests. In more general terms, this work claims that plausible immunological models should provide accurate detection mechanisms for host protection and, furthermore, that investigation on mechanisms leading to improved detection in “in silico” models can help unveil how the real immune system works.

OPEN ACCESS

Citation: Faria BF, Mostardinha P, Vistulo de Abreu F (2017) Can the Immune System Perform a t-Test? PLoS ONE 12(1): e0169464. doi:10.1371/journal.pone.0169464

Editor: Salvatore V Pizzo, Duke University School of Medicine, UNITED STATES

Received: July 12, 2016

Accepted: December 16, 2016

Published: January 3, 2017

Copyright: © 2017 Faria et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Fundação para a Ciência e a Tecnologia through grants UID/BIM/04501/2013 and SFRH/BD/79865/2011. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: This manuscript is related to a patent application, Application number, US 14/382,383, with title ‘Method and system for the detection of anomalous sequences in a digital

Background

It has long been debated whether the main function driving the adaptive immune system is related to its ability to maintain homeostasis [1, 2] or to eliminate foreign substances [3, 4]. On theoretical grounds it has been easier to build models that perform some level of self-nonsel self discrimination [5–7], even if it has been recognized that perfect self-nonsel self discrimination could be difficult to achieve [8, 9].

Instead of being focused in detecting foreignness, Niels Jerne and followers [10–12] proposed that the adaptive immune system would be concerned in maintaining homeostasis. Theoretical studies to support these ideas revolved around models of idiotypic networks, whose relevance however proved difficult to demonstrate in practice [13, 14]. After a remarkable initial growth in the 80’s, these ideas were progressively abandoned afterwards [13]. This drawback does not necessarily disprove Niels Jerne main conceptual insights, which are more general than the specific model adopted to test them. According to Jerne, immune interactions

signal'. However, this does not alter our adherence to PLOS ONE policies on sharing data and materials.

should be concerned in maintaining a regulated dynamics with itself. Foreignness could not be associated to specific antigen, but rather emerge from a perturbation in the dynamics.

The absence of a self/nonself discrimination mechanism in Jerne's conceptual model was the main source of rejection for these ideas. In an attempt to reconcile the two views, Coutinho and Varela proposed a second generation of immune networks [15]. Varela and Coutinho developed extensive numerical work and showed that lymphocytes would be arranged in a giant connected component of self-reactive elements which would be responsible for maintaining homeostasis, plus a set of disconnected peripheral and nonself reactive clones, responsible for eliminating nonself invaders. Today, these directions are still explored with new formalisms, techniques and ideas [16, 17].

Unfortunately, none of the proposed models gained indisputable acceptance being still unclear which proposal provides an effective defence mechanism [18–21]. In this respect it is important to outline the efforts played by a growing community of computing oriented researchers who have been looking into the immune system for inspiration to build better computational algorithms. Indeed, if the immune system is competent in protecting the host from invaders, new computational algorithms could use similar strategies to detect deviations from normal functioning. Considerable high quality theoretical work has been done on different formulations, such as self nonself discrimination models [6, 8], idiotypic networks [16, 17], clonal expansion models [22, 23], and models following the danger hypothesis [24]. The number of applications studied have also been impressive. It could range from fault [25, 26] and intrusion [27–29] detection to mathematical optimization [30–32] and robot path planning [33, 34]. Despite these advances, some have raised doubts on their relevance as compared to those accomplished in fields like artificial intelligence [35].

In any case, the artificial intelligence perspective has two important merits. On one side, it highlights that it is worth studying cellular processes that can encompass accurate anomaly detection, as this is likely to play a role in host protection. On the other side, it defines a research framework for testing the performance of competing theories proposing alternative mechanisms of immune protection. Indeed, our point of view is that the adaptive immune system should work as a sophisticated statistical (or data mining) detector, signalling immune responses whenever deviations from a normal state are detected. This would be analogous to the violation of the null hypothesis in statistical testing.

The Adaptive Immune System: a Modelling Perspective

Even having three different lineages (Th_1 , Th_2 and CTL), all T cells from the different subsets take part in a similar complex dynamics of cellular interactions with antigen presenting cells (APCs) (dendritic cells and macrophages) in lymph nodes. In particular, they all require stable contacts [36, 37] and two signals to become activated [38, 39]. The first signal involves interactions of T cell receptors with MHC molecules and a peptide fragment on APCs. The second signal is non-specific and typically involves CD28 and B7 molecules on T cells and APCs, respectively.

T cells play a central role in the adaptive immune system because, upon activation T cells proliferate and differentiate onto: Th_1 cells, which migrate to sites of infection, where they activate phagocytes that captured microbes with fragments like those detected by the T cell in the lymph node; Th_2 cells, which can undergo a similar type of complex dynamics with activated B cells, triggering their differentiation into plasma cells and antibody production; cytotoxic T lymphocytes (CTLs), which migrate to sites of infection and eliminate cells harbouring intracellular microbes (like viruses).

Given the central role played by T cells in initiating the immune response, the cellular frustration approach focused on modelling the interaction of APCs and T cells. The strategy has been that if plausible cellular mechanisms would be identified leading to an accurate triggering of immune responses, then this research could enlighten on the role of the different cells and signals exchanged, and what type of information the immune system senses and responds to. Indeed, it is unlikely that an inaccurate initiation of immune responses could lead to effective host protection.

To achieve flexible immunity, the immune system uses somatic recombination to build very diverse receptors, and therefore receptors are not transmitted to the organism's offspring. To accurately discriminate healthy from non-healthy cells and molecules, the adaptive immune system has to undergo a repertoire education (maturation) stage, which takes place in early life in the thymus, for T cells, and in bone marrow for B cells. Therefore, to establish viable immune protection mechanisms it is crucial to consider the two important stages in T cells' lives, the education stage where T cells are selected to recognize displayed antigen and simultaneously they are prevented from reacting against host cells, and the activation stage taking place in specialized organs in the periphery (like the lymph nodes). In both cases, cells should interact following a similar dynamics, since one stage prepares the other.

The cellular frustration framework (CFF) makes an important assumption on the scale of the fundamental processes involved in the definition of the *self*. The CFF assumes that the *self* is a systemic entity and consequently, only mechanisms processing information arising from all constituents simultaneously are likely to conveniently model the detection mechanisms involved in host protection by the adaptive immune system. This has an important modelling consequence, because it assumes that studying how individual cells interact is not enough to define the *self* information. In fact, one of the most important results in this article is the demonstration of how context dependent detection mechanisms can be built to capture systemic information, i.e., information on population properties.

To gain access to systemic information, the cellular frustration dynamics assumes that both, T cells and APCs, continuously monitor signals delivered by the cells they contact with and direct their immunological synapses towards the cell delivering the strongest signals. Experimentally it was already demonstrated that T cells can perform cellular decisions of this type (see [40] and in particular the supplementary video 4). Here we will assume that, given the extremely packed environments in the thymus and lymph nodes [41], cellular decisions are continuously taking place by both, APCs and T cells. The other assumption used is that only long contacts allow immunological synapses to mature and trigger effector functions [36, 40]. Hence, instead of being concerned with describing which cells interact with strong avidity, the cellular frustration approach [42] is concerned with which cells establish stable contacts (i.e., long-lived interactions). Furthermore, it will be the increase in the number of long long-lived interactions that will signal the degree of pathogenicity.

In this paper we will show that if one accepts the cellular frustration description of cellular interactions in the adaptive immune system, then the immune system should be capable of signalling the abnormal presentation of peptides with accuracies comparable to well-known statistical tests such as the t-test, or the KS test. This result shows how cellular interactions can aggregate information distributed over many different APCs. So far, quorum sensing is the best known biological mechanism capturing global information on a system's configuration. In contrast to quorum sensing mechanisms which are sensitive to population frequencies, cellular frustration captures information of the joint frequency of several peptides in the population, and hence it aggregates more information.

Cellular Frustration Framework as an Unstable Matching Problem

The cellular frustration approach to the adaptive immune system received inspiration from a well-known problem in computational mathematics [43], the stable marriage problem (SMP). This problem was first proposed by Gale and Shapley in 1962 [44]. Due to its relevance to market creation, work in this area was awarded the 2012 Nobel Prize in Economics. The SMP found several applications such as in organ transplant allocation [45] or management of communication networks [46].

In the stable marriage problem, researchers look for efficient algorithms matching men and women in stable pairs [44, 47]. Finding stable solutions—i.e., arrangements with only stable pairs—can be difficult because men and women all have different and complex preferences which can interfere with each other. As a result, some instances of the problem can be NP complex [47], which means that in these cases there is no known deterministic algorithm that finds a stable solution in a reasonable computational time. In any case, even for instances for which efficient algorithms exist, they are not likely to be relevant in the context of cellular immunological interactions, since they require that cells follow a precise sequence of interactions simultaneously.

The cellular frustration approach to the adaptive immune system uses the Gale and Shapley original model as a starting point. Two cell types, T cells and APCs, play the role of men and women. The crucial difference between the two models lies on their aim. The stable marriage problem looks for stable matchings because it is argued that men and women lose time when they engage in unstable matchings. By contrast the cellular frustration framework (CFF) tries to find a subset of T cells engaging in unstable (frustrated) interactions with APCs. Since effector functions require forming stable conjugations (stable pairs), in highly frustrated populations cells are rarely activated despite of their natural tendency to interact. As a result, in CFFs cellular activation is only triggered either in physiologically tolerable numbers or when the dynamics is disrupted due to changes in the information presented by APCs to T cells. Here it will be shown that, besides responses to nonself ligands (see [48]), T cells can be activated if a combination of self-ligands deviates from their typical frequencies of appearance. This is a type of anomaly response that does not require the presence of nonself ligands and for this reason we call it detection of abnormal-self.

As in the SMP, in cellular frustrated models it is assumed that APCs and T cells have preference lists, named interaction lists (ILists). Following [48], it is assumed that APCs discriminate only the presence of either one of two ligands on T cells, and consequently T cells can be grouped in two cell subtypes. Likewise, APCs can be grouped in two cell subtypes depending on which ligand they rank first. Therefore, in the model studied in this article, there are two cell types and two cell subtypes, as depicted in Fig 1. For simplicity it is also assumed that all cell subtypes have $N/2$ cells.

In the CFF it is assumed that the information displayed on cells' surfaces can be mapped onto a single ligand. APCs can display a large diversity of possible ligands. On the other hand, T cells display only 2 possible ligands, which is used to define the T cell subtype. APCs subtype is also defined using these 2 ligands as it is assumed that APCs prioritize interactions with T cells of the same subtype.

The typical complex dynamics that can emerge in the SMP arises in cellular frustrated populations as well, because of the complex organization of the ligands displayed by APCs in T cells ILists. In particular, if a T cell would rank in the top position the ligand displayed by an APC of the same cell subtype, then a conjugation between the two cells would be maximally stable. By contrast, if all ligands displayed by APCs of the opposite cell subtype would be ranked first then conjugations would be short lived because even when the T cell is conjugated

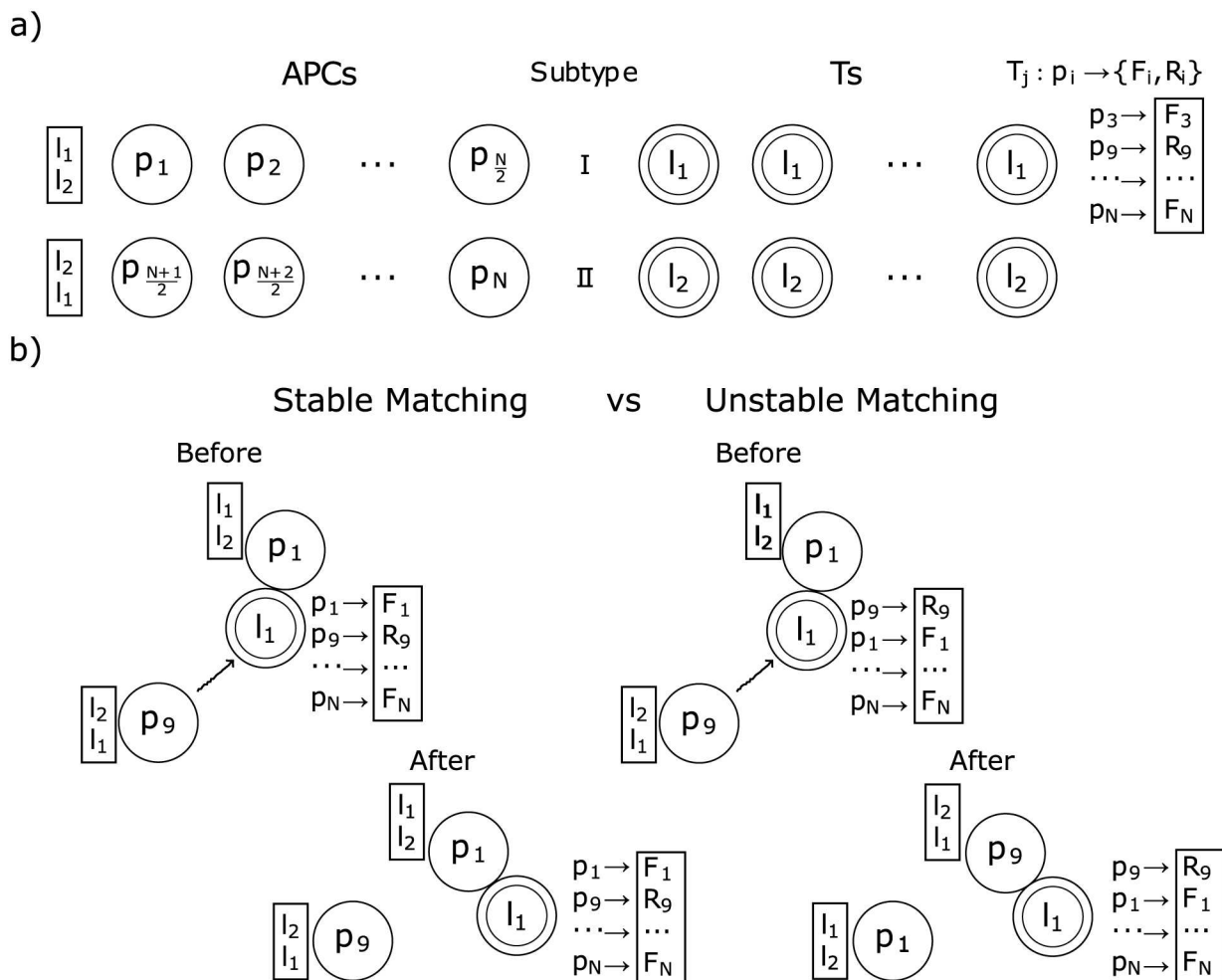


Fig 1. Cellular Frustration model used in this article. a) The model consists of $2N$ cells equally divided in two types, APCs and T cells. Each cell type is further divided in two subtypes, I and II, of the same size, and present ligands to cells of the opposite type. APCs present a diverse set of ligands p_i ($p_i \in \mathbb{R}$) while T cells present only ligands l_1 or l_2 . Ligands l_1 or l_2 determine the T cell subtype and also the APCs subtype, since APCs of subtype I (II) rank ligands l_1 (l_2) first. In this paper it is assumed that T cells map ligands p_i in two ligands (or signals), one presented frequently F_i and the other, R_i , only rarely. b) Examples of decisions taken by cells upon interaction. A new matching is formed whenever the displayed ligands are ranked higher in each other's ILists. This happens in the case on the right but not in the one on the left. In the case on the right the APC displaying the p_q ligand is said to frustrate the interaction between the APC and T cell displaying ligands p_1 and l_1 , respectively.

doi:10.1371/journal.pone.0169464.g001

to the ligand ranked in the first position, all APCs of the opposite cell subtype can destabilize the conjugation. Cells destabilizing conjugations are said to frustrate interactions. The cellular frustration framework puts a special emphasis on the importance of frustration to organize the dynamics and perform accurate intrusion detection, as discussed next.

Ordering and Detection in Cellular Frustrated Populations

The cellular frustration framework assumes that conjugation lifetimes are reliable anomaly detection indicators that can be used to trigger effector functions. This poses a problem to the immune system, namely that of measuring conjugation lifetimes using cellular contacts. In [48] it was proposed that the immune system may have solved this problem using two

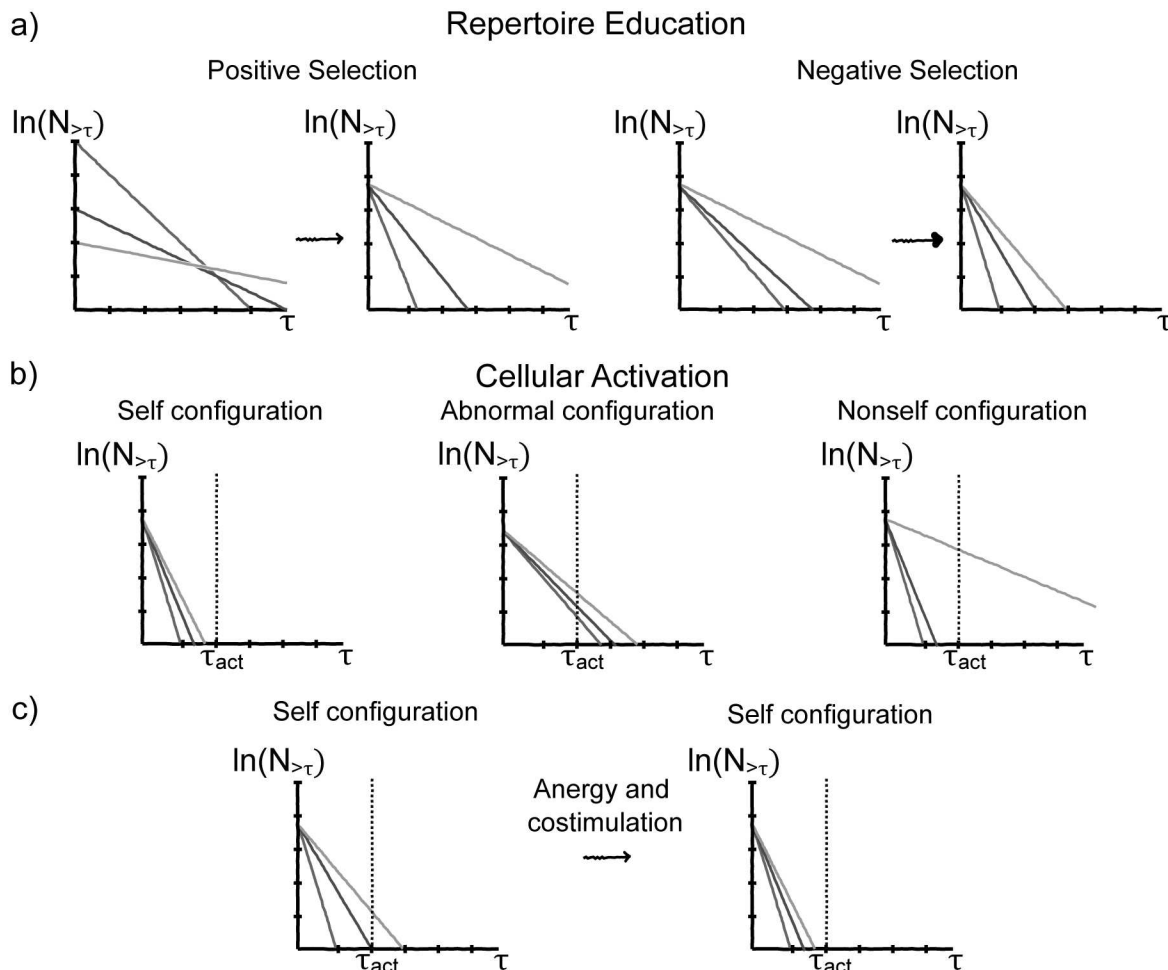


Fig 2. Important immunological mechanisms and their role according to the CFF. The number of conjugations lasting for a time τ is plotted in logarithmic scale on the vertical axis, for 3 representative cells in the population. a) (left) Positive selection homogenizes conjugation rates for all cells in the population, which is equivalent to normalizing the number of conjugations in a time interval; (right) Negative selection reduces the number of long conjugations. By combining positive and negative selection it becomes possible to access conjugation lifetimes—a reliable indicator given by the slope in the graphs—by measuring the rate of conjugations of a given duration. b) (left) Long lived conjugations, lasting longer than τ_{act} , trigger cellular activation. Self-configurations are tolerated because most conjugations are short lived ($\tau < \tau_{act}$); (middle) in the presence of abnormal self-configurations, several cells are mildly activated; (right) in the presence of nonself ligands, only a few T cells interacting with nonself ligands are strongly activated. c) By using energy and costimulation, the impact of cells that escaped education is reduced because cellular activation occurs only when several T cells produce long conjugations.

doi:10.1371/journal.pone.0169464.g002

combined mechanisms. Positive selection would homogenize the dynamics associating to all cells conjugation lifetime distributions with close normalization factors. This allows measuring conjugation lifetimes by measuring the rate of conjugations of a given duration [49] (see Fig 2a).

The other important issue concerns how frustration can be changed by ordering ILists. In [49] it was shown that conjugation lifetimes are inversely proportional to the probability of destabilizing a conjugate and consequently are related to how cells prioritize interactions. In a

stationary regime it was derived that conjugation lifetimes τ_{ij} can be calculated according to:

$$\tau_{ij} \sim \frac{1}{\sum_{kp} D_{kp|ij} \tilde{n}_{kp}} \quad (1)$$

In this expression, non-null k and i indices denote APCs, while non-null p and j indices denote T cells. Null indices are used to account for non-conjugated cells. $D_{kp|ij}$ is an integer that accounts for the number of ways kp (a conjugate or a single cell) can destabilize ij . $D_{kp|ij}$ is 0 if ij cannot be destabilized. Note that [Expression \(1\)](#) is only valid in the stationary regime. Hence, \tilde{n}_{kp} is the stationary frequency of kp conjugates or of k or p single cells (in case k or p are null, respectively).

Estimating the stationary frequencies \tilde{n}_{kp} can be difficult because APCs can present diverse information and also because their dynamical equations involve numerous feedbacks [49]. Therefore finding stationary \tilde{n}_{kp} frequencies demands self-consistent solutions which can only be found numerically. Nevertheless, this equation can still help us building a deeper understanding of mechanisms leading to accurate and sensitive immune discrimination.

First note that [Eq \(1\)](#) suggests that conjugations involving cells of the same subtype are the most stable. Indeed, in that case, only the T cell may be destabilized as the APC is already interacting with a ligand that is ranked in the top position. Therefore, by considering simply the impact of non-conjugated cells in the population, [Eq \(1\)](#) shows that the conjugation lifetime becomes at least inversely proportional to the sum of frequencies of non-conjugated APCs of the opposite cell type displaying ligands ranked above the ligand the T cell is interacting with. By contrast, if the same T cell is conjugated to an APC of the opposite cell subtype, the conjugate can be destabilized by all non-conjugated T cells of the opposite cell subtype, plus a number of non-conjugated APCs displaying ligands ranked higher in the IList than the ligand the T cell is interacting with. Thus we can conclude that long lived conjugations are mainly produced by conjugations involving cells of the same subtype. This is indeed what we obtain in numerical simulations.

Given the importance of distinguishing whether ligands are displayed by APCs of the same or of the opposite cell subtypes, we denote by LSCS, Ligands of the Same Cell Subtype and by LOCS, Ligands of the Opposite Cell Subtype.

[Expression \(1\)](#) also suggests that ILists could be organized to render the dynamics frustrated and allow accurate self-nonsel self discrimination. In [49] it was proposed that this organization could be established by the negative selection education mechanisms in the adaptive immune system. During education APCs present self-ligands. The whole set of self-ligands presented by all APCs at a given time, constitutes a configuration of the system, which changes from time to time. Negative selection eliminates all T cells interacting with high affinity with ligands displayed by APCs and replaces them by new incoming T cells (with random ILists). Within the CFF, high affinity interactions are measured by conjugation lifetimes and consequently, T cells are eliminated because they establish stable conjugations with APCs of the same cell subtype displaying ligands that are ranked in the T cell IList top positions. Therefore and except for a small fraction, only T cells without ligands displayed by APCs of the same cell subtype on top positions of their IList can survive negative selection. This is true for all ligands frequently displayed by APCs. However, nonself ligands were not displayed by APCs. Consequently they cannot have made any impact on the education process and therefore, T cells surviving negative selection can still have them ranked on top positions. This implies that if nonself ligands are displayed by APCs in the periphery, many T cells may establish stable conjugations with these APCs.

In [48] it was argued that the CFF offers an integrated and consistent understanding of how the immune system performs immune detection. In particular, it was shown that costimulation and anergy improve considerably the accuracy of nonself discrimination by reducing the impact of errors in negative selection. This agrees with the interpretation of experimental data [50]. The CFF proposes that anergy and costimulation guarantee that APCs are only activated if multiple different T cells – and not just a single cell which could have escaped proper negative selection – establish long conjugations with an APC. Anergy is also important to force the immune system to test a wide range of different T cell receptors.

Costimulation and anergy are two of the three signals [38] required to mount an immune reaction. Within the CFF the third signal, which is usually associated to cytokines delivered to the medium [51], naturally arises from the fact that immune reactions should only be mounted once the sum of activation signals over all cells exceeds a threshold.

Fig 2 summarizes the main ideas and consequences that make cellular frustration a consistent framework. In the next section we will discuss how cellular frustrated systems (CFSs) detect signals of anomalies not arising from nonself ligands.

Abnormal Self detection in Cellular Frustrated Systems

Consider a conjugate with cells of the same cell subtype in a frustrated population. Then, as discussed in the previous section, the conjugate can only be destabilized by non-conjugated APCs displaying LOCS that are ranked in top positions in the T cell IList. If the number of these LOCS diminishes, the denominator in Eq (1) decreases, increasing the average conjugation lifetimes. Since in the CFF, long-lived conjugations trigger effector functions, this suggests that CFSs should be capable of detecting other signs of intrusion besides the presence of nonself ligands. In fact, we will show that CFSs can detect two other types of perturbations. One is the increase in the number of ligands that are only rarely presented. These ligands are not nonself, because they appear in self-configurations. However, they appear only rarely and in this work they can be seen as the ligands appearing on the tails of distribution functions.

The other type of perturbation CFSs can react to occurs when uncommon combinations of frequent ligands are absent. This is a much more complex type of information as it captures correlations in presentation patterns. This detection mode goes well beyond single ligand statistics.

In this work it is assumed that T cells sense only two types of signals from each APC. From a modelling perspective this is equivalent to assume that each T cell maps the information displayed by each APC onto only two different ligands, i.e., as if only each APC would present one of two possible ligands. According to our results, the best immune protection is achieved when one of these ligands is perceived frequently and the other appears only rarely. This perspective agrees with that of several authors [39, 52–54]. Therefore, from here-on we refer to the information perceived as delivered by frequent or rare ligands. Note however, that this mapping differs from one T cell to the next, and that different APCs display different information and therefore their information is mapped onto different ligands. Note that different metrics can be used to establish this mapping, depending on the modeling taken to establish how T cells read the information presented by APCs. For instance, some models could focus on the interaction of the T cell receptor and the peptide-MHC complex, while others on avidity effects (see [55]).

An intuitive picture can then summarize how abnormal-self detection is achieved in CFSs. First, negative selection selects which sub-sets of ligands can be ranked in a small fraction of T cells ILists, on top positions, to minimize the frequency of long lived conjugations. Since negative selection uses a common and progressively adjusted lifetime threshold to eliminate T cells,

all T cells adjust together the number of ligands kept under surveillance and in this way correlate their responses. As a result, after negative selection each T cell continuously surveys the presence of the set of frequent ligands it has ranked on top positions in its IList and displayed by APCs of the opposite cell subtype. If any of them is absent, the probability of establishing long conjugations with APCs of the same cell subtype increases. The amplitude of these individual cell responses is larger, the larger the number of absent frequent ligands. Using the long conjugation lifetimes observed in self-configurations, cellular activation thresholds can be defined.

The immune system as a whole is activated when the sum of individual cells activation signals (e.g., in the form of cytokines concentration) exceeds a threshold. This collective signaling can be related to the third signal required for the immune system activation, and it is stronger, the larger the number of activated cells. Hence, abnormal self-detection is triggered depending on the number of frequent ligands missing and on the number of T cells having them ranked in top positions.

Results

Two types of results are presented next. First, the several detection mechanisms used by CFSs will be identified. This is best addressed using a specifically designed case study. Afterwards it will be shown that CFSs can achieve state of art anomaly detection performances in more general settings.

DinBs: a case study with data presented in blocks

Consider a simplified version of the CFS model, in which N_B APCs belonging to first cell subtype either present a frequent or a rare ligand. In this example, it is assumed that all APCs present different ligands and all T cells use the same criteria to establish whether APCs display a rare or a frequent ligand ($N_B = N/4$). This is a special case of the more general model considered in this article, in which all T cells perceive differently the information presented by the several APCs. Since, in the models presented here, T cells map this information onto only 2 signals (or ligands), a frequently or a rarely displayed, this is also equivalent to assume that APCs present only either a frequent or a rare ligand (Fig 3).

In self-configurations either N_r^{left} rare ligands are displayed by the first N_B APCs of the first cell subtype, or N_r^{right} are displayed by the remaining APCs of the same subtype. The total number of rare ligands displayed in a configuration s is then $N_r(s) = N_r^{left} = N_r^{right} = N_r$.

Discrimination of two types of abnormal self-configurations will be tested. In the first case the number of rare ligands is increased: $N_r^{left} = N_r^{right} = N_r + \Delta N_r$. In the second case, the same number of rare ligands, N_r , is displayed as in self-configurations. However, now half of them are presented by cells from the last $N_B/2$ cells from the first block and the other half by cells from the first $N_B/2$ cells from the second block of the first cell subtype. A representation of this data presentation scheme is shown in S1 Fig. An extension of this model considering that rare ligands could be presented by cells belonging to both cells subtypes would lead to similar results and therefore will not be considered here.

Maximally frustrated populations for CFSs with DinBs. Our initial analyses consider T cell populations with partially ordered ILists. These lists maximize frustration for all cells simultaneously, minimizing the longest conjugation lifetimes. To maximize frustration in the SCFSs with DinBs, ILists are organized according to the following strategy. First rank in top positions N_{LOCS}^{top} frequent LOCSs; afterwards rank rare LSCSs; then rank frequent LSCSs; then rank rare LOCSs and finally rank the last frequent LOCSs. The specific way ligands are ranked in different T cells ILists also follows a specific order (see S2 Fig).

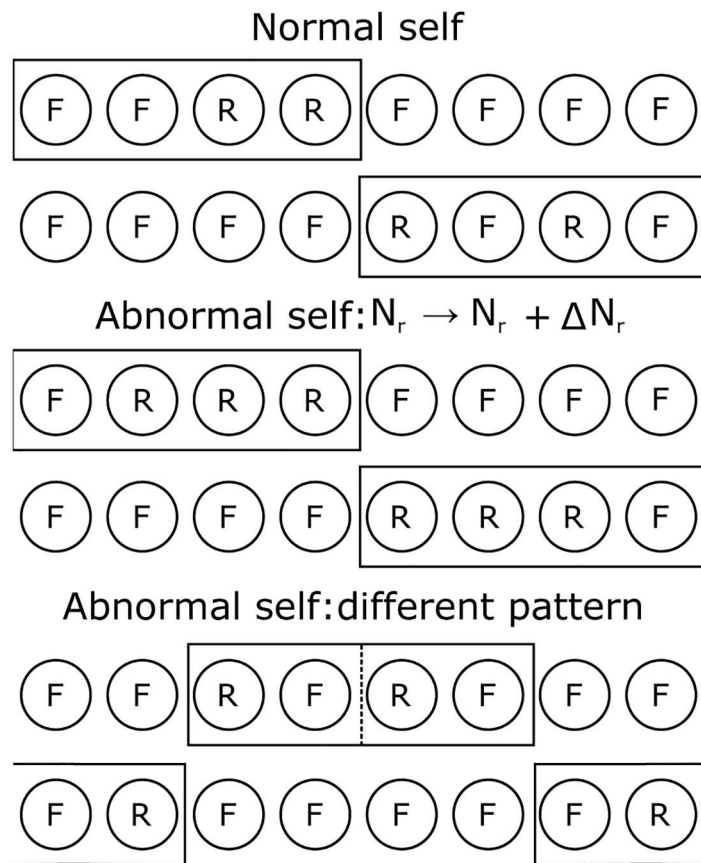


Fig 3. Illustration of the configurations displayed by APCs in the Data in Blocks (DinBs) case study. In circles and on each row are represented APCs of subtype I. Each APC can only present two ligands, represented either by an F or an R depending on whether they are frequently or rarely displayed. Different APCs display different ligands. Rectangular boxes delimit APCs that can randomly assigned to display rare ligands. For configurations with rectangular boxes with a dashed line, half of the number of rare ligands is displayed by APCs on one side of the box, and the other, on the other side. Configurations on the first four rows share the same presentation pattern but differ on the number of rare ligands displayed. Configurations corresponding to the last two rows, display sets of rare ligands that have never been displayed together when normal self-configurations were presented during education (maturation).

doi:10.1371/journal.pone.0169464.g003

Mechanisms increasing the number of long lived conjugations. The number of stable conjugations can be increased in two ways: by increasing the probability of establishing stable conjugations by individual cells, or by increasing the number of cells establishing stable conjugations.

To establish long conjugation rates, T cells have to establish stable interactions with LSCS. This happens when frequent LOCS ranked in top positions in T cells ILists are absent. Indeed, according to Eq (1), the largest conjugation rates increase with $\tau^{-1} \propto \sum \tilde{n}_{k0}$, where the sum over k runs over all APCs displaying LOCS on top positions in T cells ILists. One can then expect that the number of long lived conjugations established by a T cell with index i should be equal to $n_i^0(\tau_{act})/n_i^0(0) = \exp(-\tau^{-1}\tau_{act}) = \exp(-\gamma N_{i,LOCS}^{top}\tau_{act})$, where $N_{i,LOCS}^{top}$ denotes the number of LOCS on top positions in the i^{th} T cell IList, γ a proportionality constant and $n_i^0(0)$ a normalization constant. If $N_{i,LOCS}^{top}$ is decreased to $N_{i,LOCS}^{top} - \Delta N_{i,LOCS}^{top}$ then $n_i(\tau_{act})/n_i(0) = n_i^0(\tau_{act}) \exp(-\gamma \Delta N_{i,LOCS}^{top}\tau_{act})$. This is indeed confirmed in the results obtained

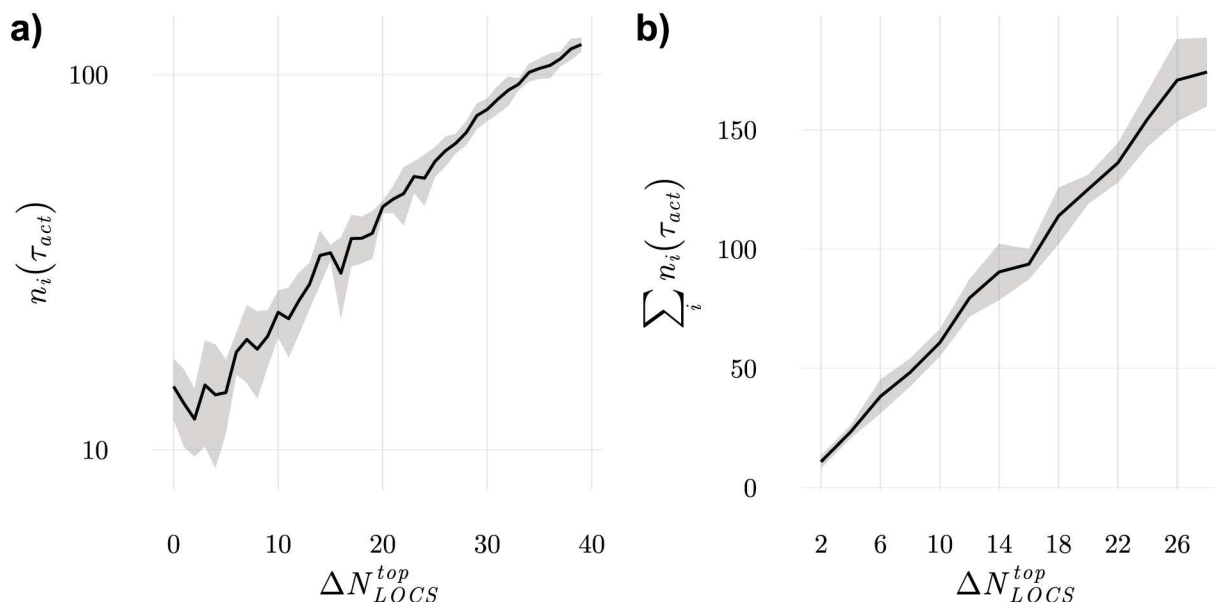


Fig 4. Cellular responses when ILists are specifically modified to analyse the impact of missing frequent ligands. a) the number of long lived conjugations grows exponentially when a growing number of frequent LOCS are removed from top positions in a T cell IList; b) the number of long lived conjugations grows linearly when a growing number of cells has a couple of frequent LOCS removed from their ILists top positions. In these results 100 realizations of systems with 96 cells were used.

doi:10.1371/journal.pone.0169464.g004

in numerical experiments shown in Fig 4. In these experiments, all frequent LOCS are ranked on top positions in T cells ILists, except for those in one cell in which an increasing number of these ligands is moved to the lowest positions. The exponential dependency of the number of long lived conjugations on the number of removed ligands, ΔN_{LOCS}^{top} , can then verified.

From these results we can conclude that T cells work as multiple correlation function evaluators responding whenever combinations of LOCS are absent. On this respect it should remarked that T cells are extremely efficient at performing this task since they evaluate multiple combinations at once. Another conclusion that can be drawn is that the established long conjugations are specific with respect to which LOCS are absent, but are not specific relatively to which LSCs are actually producing the long lived conjugations. Indeed, changing the ranking of LSCs in T cells ILists does not change the probability of establishing long lived conjugations. This is in strike contrast with nonself discrimination [48] in which case the APC that triggers the response displays the ligand recognized as nonself.

Instead of requiring that T cells are heavily perturbed, it is also possible to obtain the same increase in the number of long lived conjugations by increasing the number of T cells that engage, even if only mildly, in stable conjugations in the whole population. Then, in a first approximation equation (Eq (1)) predicts that if a number of T cell ILists is mildly perturbed, then long lived conjugation rates should increase linearly to their number. Likewise, the total number of long lived conjugations established in the whole population should increase linearly with the number of perturbed T cells. In Fig 4b we show that indeed, this is the case, irrespectively to which ligands disappear. This result shows that activation signals arising from multiple cells can add up to create proportionally reliable signals.

T cells detect increments on the number of displayed rare ligands. Now we show that T cells can detect deviations from normal presentations that are characterized by the presentation of a larger number of rare ligands than in self-configurations.

Denote by $\tau_{ij}^0(s)$ the conjugation lifetime performed by APC $i \in \mathcal{A}$ (\mathcal{A} being the set of all APCs) and T cell $j \in \mathcal{T}$ (\mathcal{T} being the set of all T cells) when a configuration $s \in \mathcal{S}^0$ is presented (\mathcal{S}^0 being the set of all configurations of self-ligands presented in the thymus). Denote by $\tau_{ij}(s)$, the corresponding conjugation lifetime for configurations presented in the periphery. Then, in general, $s \in \mathcal{S}$ with $\mathcal{S}^0 \subset \mathcal{S}$.

The first important observation is that the T cell population can discriminate, in principle, the presence of any rare ligand. To see this, consider first, that in self-configurations frequent LOCS had never been absent. Then, when the rare ligand is introduced in the maximally frustrated populations defined above with all frequent LOCS on top positions in T cells ILists (and all rare LOCS on the bottom), all T cells increase their largest conjugation lifetimes (with LSCs) since the number of frequent LOCS is reduced when a rare ligand appears. Consequently, $\max_{i \in \mathcal{A}}(\tau_{ij}(s')) > \max_{i \in \mathcal{A}}(\tau_{ij}^0(s))$, $\forall s' \in \mathcal{S}$, $\forall s \in \mathcal{S}^0$, $\forall j \in \mathcal{T}$. In principle, the same result should be obtained when only subsets of LOCS are on top positions in T cells ILists, provided all frequent LOCS appear in top positions in T cells ILists.

This was indeed confirmed in numerical simulations that counted the number of conjugations $n_j(\tau_{act})$, lasting τ_{act} instants in a fixed time interval and established by a T cell with index j . In any configuration in which a rare ligand is introduced, there are always T cells for which $n_j(\tau_{act}) > \max(n_j^0(\tau_{act}))$, where $n_j^0(\tau_{act})$ is the number of long lived conjugations in normal configurations with no rare ligands displayed. Therefore discrimination is perfect in this extreme case.

The previous result requires that the number of rare ligands displayed in self-configurations is zero. This, however, is a serious drawback since it does not allow discriminating disturbances that are not linked to an increase in the number of rare ligands, but are due to different patterns of absence of frequent ligands (or, equivalently, different patterns of presentation of rare ligands).

To discriminate this type of disturbances it is crucial that frequent ligands are absent in self-configurations. In this case, however, an increase in the number of absent frequent ligands is not forcefully discriminated. This can be explained in a simple way by considering that one rare ligand is presented in self-configurations but, in the periphery, two rare ligands are presented instead. To tolerate the absence of one frequent ligand in any self-configurations, T cells should require that two LOCS ranked in top positions are absent in order to become activated. This, however, is not forcefully achieved if ILists only have subsets of frequent LOCS on top positions. In this case, triggering the system depends on the probability of activating at least one T cell. As we will show in the next section, this is the case of practical relevance, and therefore it should be studied in more detail.

Consider a maximally frustrated population with DinBs and with rare ligands presented only by APCs of one subtype in one of the two blocks. From the previous analysis, it follows that when the number of presented rare ligands is small, T cell activation should require the absence of two or more frequent LOCS ranked in top positions. The probability that f frequent LOCS are absent from top positions in T cells ILists is given by:

$$P(f) = \frac{\binom{N_{LOCS}^{top}/2}{f} \binom{N_L/2 - N_{LOCS}^{top}/2}{N_r - f}}{\binom{N_L/2}{N_r}} \quad (2)$$

where $N_{LOCS}^{top}/2$ is the number of frequent LOCS from one block ranked in top positions, N_r is the number of rare ligands presented in the configuration and $N_L/2$ is the number of ligands

from one block presented by APCs of subtype 1. The probability of activating a T cell is then $P_{act} = 1 - P(0) - P(1)$. The activation of the whole system requires that a pre-defined number of T cells to be activated. This number depends on the false positive rate the system can safely accommodate. If one assumes that this false positive rate is denoted by α , then the number of T cells that must be activated to activate the whole system, N_a , is the solution of the following equation:

$$\sum_{n=0}^{N_a} \binom{N_L/2}{n} P_{act}^n (1 - P_{act})^{N_L/2-n} = 1 - \alpha \quad (3)$$

Although the number of T cells of each subtype is N , by construction only half T cells have different IList in maximally frustrated systems for DinBs. The threshold value can be found by solving the fixed point recursive equation:

$$N_{a,n+1} = N_{a,n} - \left\lfloor \lambda \sum_{n=0}^{N_a} \binom{N_L/2}{n} P_{act}^n (1 - P_{act})^{N_L/2-n} + \alpha - 1 \right\rfloor \quad (4)$$

where λ is a parameter chosen to guarantee convergence (typically $\lambda = 0.1$) and $\lfloor \cdot \rfloor$ denotes the floor operation. The asymptotic solution of this equation gives the (threshold) number of cells above which the system is activated as a whole: $N_a^{thres} = N_{a,\infty}$. To evaluate the discrimination power when the number of rare ligands is increased, the probability of activating the whole system is calculated from:

$$\Xi(N_r^0 + \delta N_r) = \sum_{n=N_a^{thres}}^{N_L/2} \binom{N_L/2}{n} (P_{act}^*)^n (1 - P_{act}^*)^{N_L/2-n} \quad (5)$$

where here P_{act}^* is the probability of activating a T cell when $N_r = N_r^0 + \delta N_r$, N_r^0 being the number of rare ligands presented in self-configurations. The results are presented in Fig 5, for $N_r^0 = 1, 2$ and when δN_r is successively increased until full discrimination is achieved. Results obtained using the agent based numerical simulations with maximally frustrated populations are also depicted showing a good agreement. Furthermore, results from numerical simulations using the complete cellular frustration algorithm, with education of T cells ILists (discussed later in this article) are also shown, in Fig 6. The agreement among these results, confirm our interpretation of the mechanisms at play.

Two main conclusions can be drawn from these results. First, increasing successively the number of rare ligands increases the probability of discriminating the perturbation according to a logistic type of saturation growth curve. Indeed, consider the addition of a rare ligand, $N_r \rightarrow N_r + 1$. Then, in a fraction of configurations the whole system is activated. If another ligand is introduced afterwards, a fraction of the formerly non activated configurations will have a similar probability of activating the system. If one assumes that this probability is roughly constant, then full discrimination is achieved and the logistic type of saturation growth curve displayed in Fig 5 is explained. More importantly, this result shows that the addition of an increasing number of rare ligands can eventually be always discriminated.

The second conclusion that is worth discussing is that the larger N_r^0 the harder it becomes to detect the addition of a rare ligand when N_r^0 is increased to $N_r^0 + 1$ ($N_r^0 \rightarrow N_r^0 + 1$). Understanding this effect is important because it highlights the trade-off between tolerance and immunity. Consider the set of all self-configurations with N_r^0 rare LOCS (see S3 Fig). A large fraction of these configurations, $1 - \alpha$, has to be tolerated, while the remaining produce false positive activations.

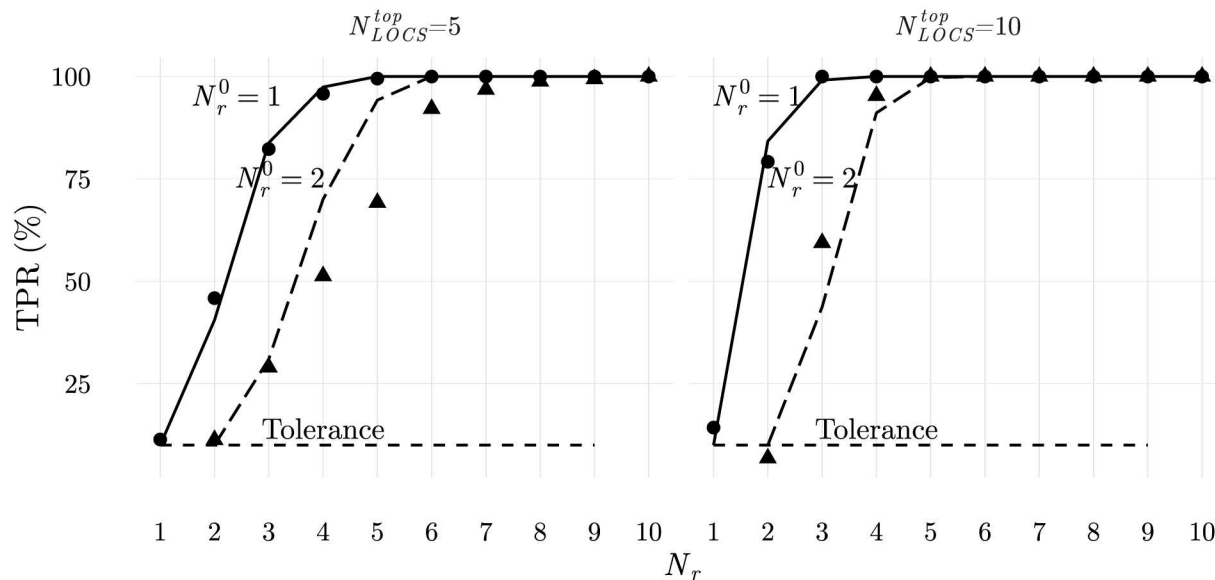


Fig 5. Average true positive rate when configurations with N_r rare ligands are presented to populations calibrated with $N_r^0 = 1$ or $N_r^0 = 2$ rare ligands. Results in dots and triangles are from numerical simulations where T cells have perfectly ordered ILists with $N_{LOCS}^{top} = 5$ (left) or $N_{LOCS}^{top} = 10$ (right) frequent LOCS in top positions from each block. Simulations used 96 cells of each type and 100 realizations. Solid and dashed lines are the predictions from the theoretical arguments described in the text. A false positive rate of 10% was assumed.

doi:10.1371/journal.pone.0169464.g005

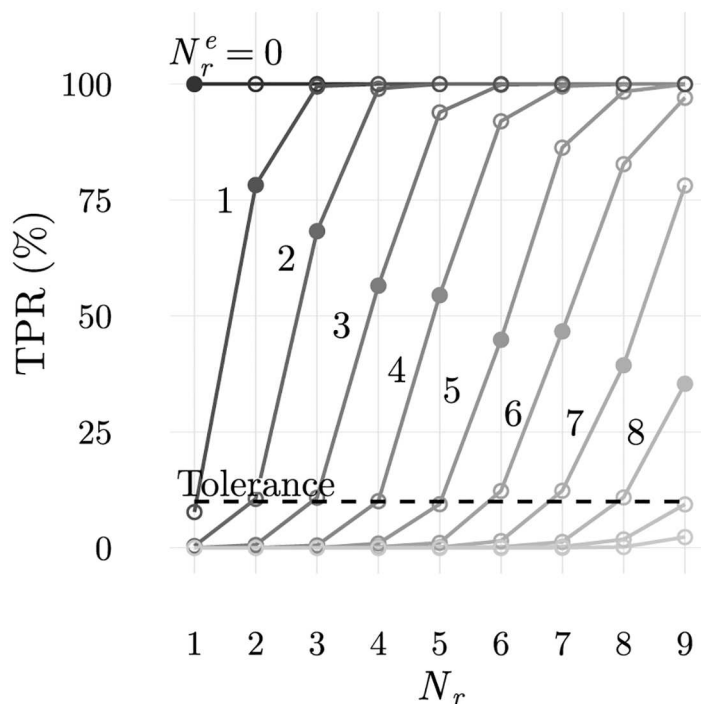


Fig 6. Average true positive rate when configurations with N_r rare ligands are presented to populations obtained by repertoire education and calibrated with several N_r^0 rare ligands. Highlighted with solid dots are results obtained when a single rare ligand is added to normal configurations. Larger numbers of rare ligands in normal configurations introduce noise reducing the capacity to detect perturbations.

doi:10.1371/journal.pone.0169464.g006

Next consider the set of configurations with $N_r^0 + 1$ rare LOCS. To each configuration with N_r^0 rare ligands, $N/2 - N_r^0$ configurations ($N/2$ being the number of subtype I APCs) can be constructed with $N_r^0 + 1$ rare ligands. Clearly, the fraction α of activated configurations with N_r^0 rare ligands, will also be activated when an extra rare ligand is added. Furthermore, a set of other configurations will also become activated, totally a fraction δ of activated configurations, corresponding to the fraction of true positives. Assume that $\alpha < \delta$, i.e., there is discrimination.

Next consider what happens if the number of rare ligands in self-configurations is $N_r^0 + 1$, and the population should detect the variation $N_r^0 + 1 \rightarrow N_r^0 + 2$. From the fraction δ of activated configurations in the previous case, only a fraction α of the most activated configurations will activate the system, i.e., activation thresholds are now more demanding. Therefore, the set of configurations with $N_r^0 + 1$ rare ligands can be divided in 3 subsets: the most activated corresponding to the false positives when $N_r^0 + 1 \rightarrow N_r^0 + 2$; the least activated, corresponding to false negatives when $N_r^0 \rightarrow N_r^0 + 1$ (configurations not activated and with $N_r^0 + 1$ rare ligands); and the remaining activated configurations when $N_r^0 \rightarrow N_r^0 + 1$. From this analysis it becomes clear that the discrimination power decreases when $N_r^0 + 1 \rightarrow N_r^0 + 2$ relatively to $N_r^0 \rightarrow N_r^0 + 1$, because the probability of creating activated configurations with $N_r^0 + 2$ rare ligands from the least activated configurations with $N_r^0 + 1$ rare ligands is smaller than the probability of generating these configurations with the intermediately activated configurations with $N_r^0 + 1$ rare ligands. Therefore, very general arguments explain the behaviour observed in the results in Fig 6.

In general, those configurations that previously activated the system, are now more likely to trigger the immune system when $N_r^0 + 2$ rare LOCS are displayed. This is because these configurations already have a larger number of activated cells, and therefore the number of T cells that should also become activated is smaller, or because there is a higher probability of further removing a frequent LOCS from the top of an activated T cell IList. However now, since thresholds are more demanding, only for a fraction of configurations the removed frequent LOCS can activate the immune system and consequently it should be expected a smaller discrimination power.

In simpler terms it could be stated that introduction of rare LOCS introduces noise which makes discrimination harder. However, in the next section it will be shown that this allows detection of other types of perturbations.

The previous mathematical approach assumed that populations displayed a small number of rare ligands in the thymus. In this limit, activation was controlled by the number of T cells sensing the absence of 2 frequent LOCS in tops positions. For self-configurations with a larger number of absent frequent LOCS, the threshold controlling the system's activation may impose that a combination of T cells sense the absence of a different number of frequent LOCS (e.g., a fraction sense the absence of 2 while another fraction the absence of 3 frequent LOCS) since the system's response aggregates the responses of all individual cells. The CFF proposes a set of immunologically plausible mechanisms to make this adaptive selection of criteria automatic. Clearly, the immune system, and agent based simulations in particular, have the advantage of being able to tune thresholds to adapt to complex presentation patterns.

T cells can sense contextual information. A more challenging detection capability consists in detecting disturbances that are not due to an increase in the number of rare ligands. Instead, it is the combination of ligands that determines whether a configuration is normal or abnormal. In this case the information is contextual. Within the CFF, T cells can still respond to this type of disturbances since responses depend on the combinations of absent frequent LOCS. However, a number of conditions have to be met. Firstly, the mapping into rare and frequent ligands should make rare ligands sufficiently frequent in order to provide

information of presentation patterns. Secondly, frequent LOCS should not all be ranked on top positions. If this would be the case, then the number of absent frequent LOCS would not change as their overall number does not necessarily increase. Finally, and for the same reason, the system's response should only receive contributions from cells sensing the absence of frequent LOCS as if all cells contributed their responses could cancel out. This naturally shows why single cell activation thresholds are needed.

Now we apply the previous mathematical analysis on maximally frustrated populations in the DinBs case study. To guarantee T cell's tolerance towards the two types of normal configurations, T cells ILists should have on top positions the same number of frequent LOCS displayed by APCs of the two blocks. In this way, it is guaranteed that any T cell has at least $N_{LOCS}^{top}/2$ frequent LOCS on top positions. This number increases to $N_{LOCS}^{top} - N_R$ if $N_r < N_{LOCS}^{top}/2$.

Consider a specific example where $N_{LOCS}^{top} = N_L/4$, i.e., each T cell has on top positions half of the whole set of frequent LOCS of each block of data. In abnormal configurations, the same number of rare ligands is presented, but now half is presented on a block on the right and the other half on a block on the left (see Fig 3). The minimum number of frequent LOCS in ILists top positions is $N_{LOCS}^{top} - N_R$, but now this value has 0 as lower bound – see S4 Fig. Therefore, some cells sense the absence of a considerable larger number of frequent LOCS and will respond to the contextual change.

Consider that only two rare ligands are presented, both in normal and abnormal configurations. In this case, all steps taken before to find the activation threshold, N_a^{thres} are still valid since normal configurations are the same. What differs is the population response to abnormal configurations. Since T cells should not have all frequent LOCS on top positions in their ILists, they sense the presence of frequent LOCS differently.

By construction, when $N_{LOCS}^{top} = N_L/4$, there are 2 ILists (identical) with N_c frequent LOCS in the central block (half on the left and the other half on the right) and $N_{LOCS}^{top}/2 - N_c$ frequent LOCS outside. Since one rare LOCS is presented on each side, the probability that 1 frequent LOCS is missing from top positions in an IList is:

$$p_1 = \frac{\binom{N_c/2}{1} \binom{N_L/4 - N_c/2}{0}}{\binom{N_L/4}{1}} = 2 \frac{N_c}{N_L} \quad (6)$$

The probability that a T cell with N_c frequent LOCS in the central block, misses 2 frequent LOCS (one on each side) is:

$$P(2, N_c) = p_1^2 = 4 \frac{N_c^2}{N_L^2} \quad (7)$$

Finding the probability of having more than N_a^{thres} cells with 2 frequent LOCS missing is harder to calculate now than in the previous case. It amounts to calculate a convolution with N_a^{thres} probability distributions. We computed this numerically by using a Monte Carlo approach. For a system with 48 APCs of each subtype ($N_L = N/2 = 48$) and when $N_{LOCS}^{top} = 24$, we obtained $N_a^{thres} = 9$ and $\Xi_{context}(N_r = 2) = 13.5\%$ for $\alpha = 0.1$.

This agrees with the average true positive rate of 14.1% obtained after simulating 100 populations using the cellular frustration dynamics on populations with maximally frustrated ILists and shown in Fig 7. There, detection rates obtained when the number of frequent LOCS in top positions is varied are also shown.

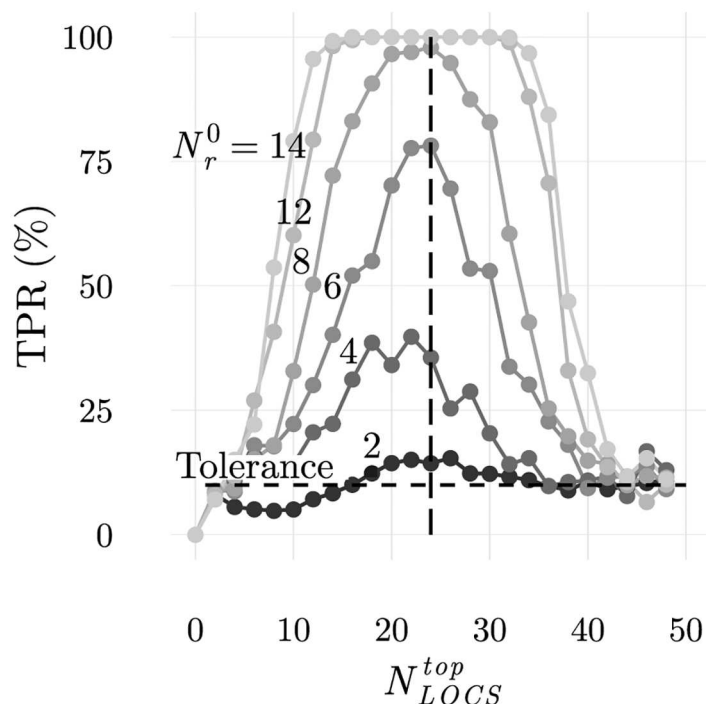


Fig 7. Average true positive rate for simulations for context dependent (abnormal self) discrimination. In the simulations considered for these results, both normal and abnormal configurations displayed the same number of rare ligands, although in different patterns as shown in Fig 3. Averages accounted 100 realizations and a false positive rate of 10%. Populations with 96 T cells and partially ordered ILists with N_{LOCS}^{top} were used. From these results it is clear that the best discrimination is achieved for partially ordered ILists that maximize the number of (potentially) absent frequent LOCS in top positions. In the present case this number is 24, corresponding to the size of an ILists with all LOCS from the block presented in the abnormal self configuration. It is also clear that the larger the number of absent frequent LOCS, the higher is the discrimination.

doi:10.1371/journal.pone.0169464.g007

Two important conclusions can be drawn from these results. Firstly, the number of frequent LOCS that should be ranked in top positions to maximize discrimination depend on the data presented. In this case rare ligands were presented in blocks with 24 APCs of the first sub-type, and the maximum number of different ligands between normal and abnormal configurations was 24. As a result, the number of frequent LOCS that should be ranked in top positions to maximize discrimination was also 24. This shows that the immune system should find ways to automatically adapt to the information presented, ranking only an adequate number of frequent LOCS in top positions. This, of course contrasts with what was required to obtain the best results if the immune system was only concerned with detecting an increase in the number of displayed rare ligands. Therefore a trade-off exists as highlighted by these results.

Secondly, even if there are 24 frequent LOCS ranked in top positions (results along the dashed vertical line in Fig 7, the number of absent ligands is important to achieve the highest discrimination. In Fig 7, a minimum number, higher than 16, is required to obtain a 100% true positive rate. Increasing beyond this number, does not improve discrimination of contextual information, but it could be prejudicial for discriminating increments in the number of displayed rare ligands.

In Fig 8 the true positive rate obtained with populations of T cells educated following negative selection, are also shown. These results show that the mechanisms of negative selection,

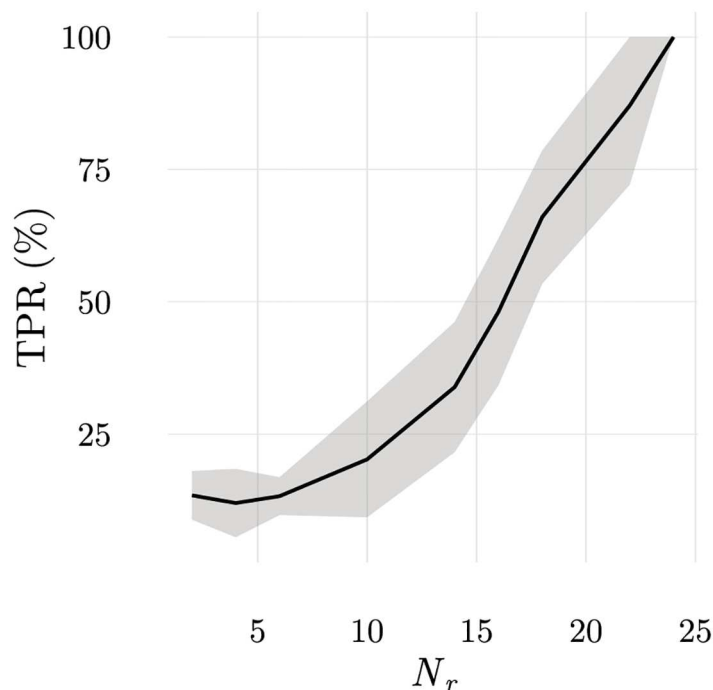


Fig 8. Abnormal self discrimination for populations with educated ILists when the number of rare ligands displayed is varied. Increasing the number of rare ligands in self and abnormal self-configurations increases discrimination. This result agrees with the theoretical arguments developed under the more restrictive conditions of Fig 7. In particular, it shows that context information can be perfectly discriminated even when ILists are ordered by negative selection.

doi:10.1371/journal.pone.0169464.g008

which will be analysed more thoroughly next, also demonstrate that discrimination of contextual information is possible and improves with the number of rare ligands displayed. Results obtained with negatively selected populations lead however to poorer performances, which is expectable since maximally frustrated populations had been designed to obtain the best performances.

Negative selection is heterogeneous and selects T cells with the largest number of LOCS on top positions. The previous analysis provided two important results. The first was that contextual discrimination is possible and, in certain cases, even perfect. The second was that ILists should have on top positions a balanced number of ligands of each block so that there is always a minimum number of frequent LOCS present in any self-configuration. The important issue that should be addressed next consists in understanding how ILists organization and the number of LOCS in top positions can be chosen autonomously. In the adaptive immune system this is achieved in the thymus through repertoire education mechanisms. Within the CFF, negative selection operates by eliminating T cells producing the longest conjugation lifetimes. Since the longest conjugations occur for those cells sensing the absence of the largest number of frequent LOCS, negative selection should select T cells that rarely have less than a (threshold) number of frequent LOCS absent. When applied to the DinBs case study this is indeed what is observed, as shown in Fig 9. Therefore, negative selection produces the ordering required for contextual discrimination, as discussed in the previous section.

Fig 10 shows two other important points. Firstly, when the maximum conjugation lifetime decreases, the average position of the LSCS ranked in the highest position in T cell ILists

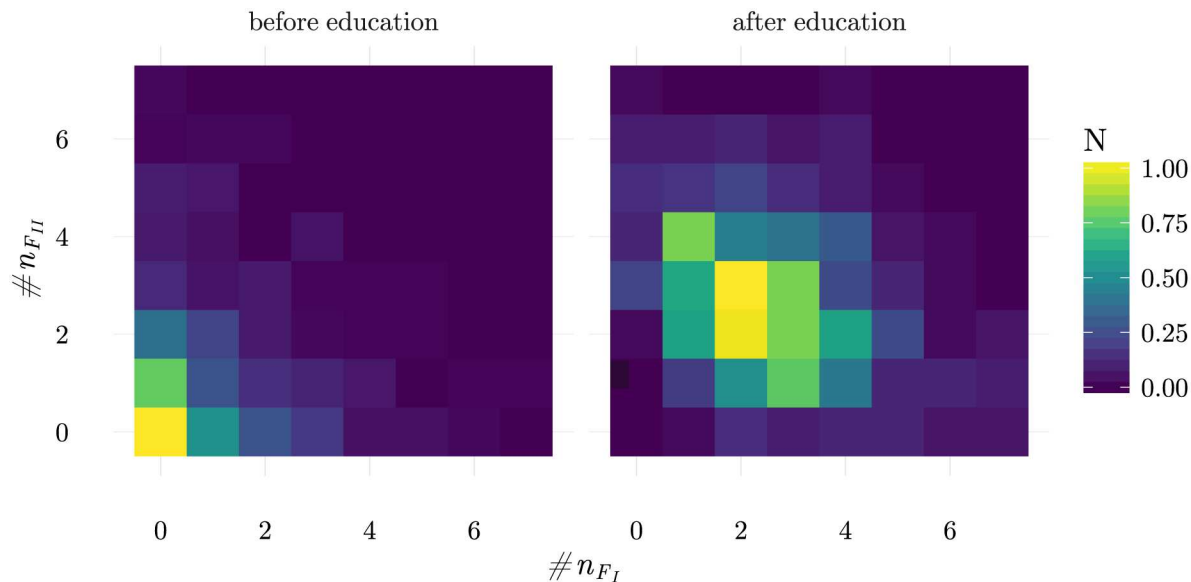


Fig 9. Frequency of the number of frequent LOCS of each block (I and II) on ILists above the highest ranked LSCS, before and after education (maturation). These results were obtained considering self-configurations with $N_r = 6$ (see [Methods](#) for other remaining simulation parameters). These results show that repertoire education balances the number of frequent LOCS of each block on ILists top positions, guaranteeing that a minimum number of frequent LOCS is always present in self-configurations.

doi:10.1371/journal.pone.0169464.g009

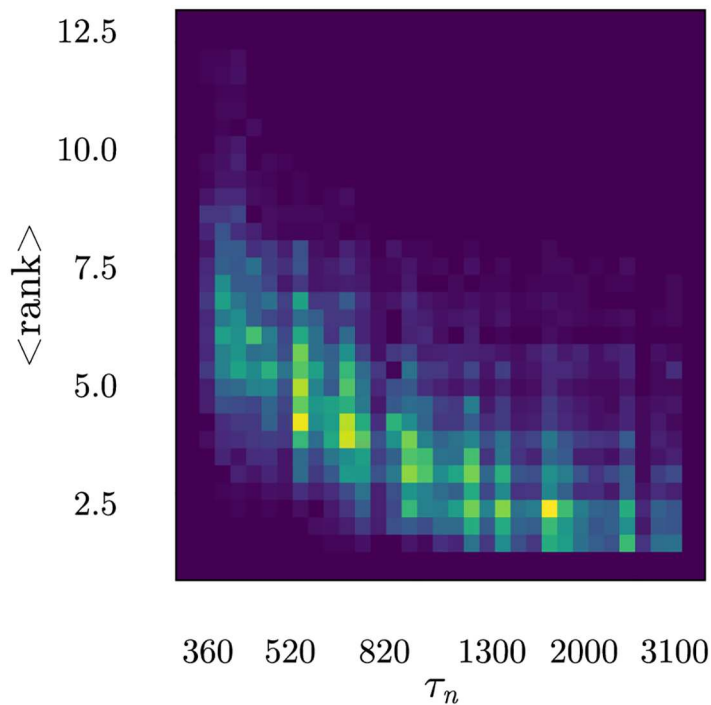


Fig 10. Two dimensional histogram for the mean rank of the first ranked LSCS in each IList as a function of the maximum conjugation lifetime τ_n in the population along repertoire education. It is clear from these results that LSCSs occupy progressively positions more on the bottom of ILists. For this figure 10 repetitions of the education of 120 populations with 96 T cells were considered. Self-configurations had $N_r = 6$.

doi:10.1371/journal.pone.0169464.g010

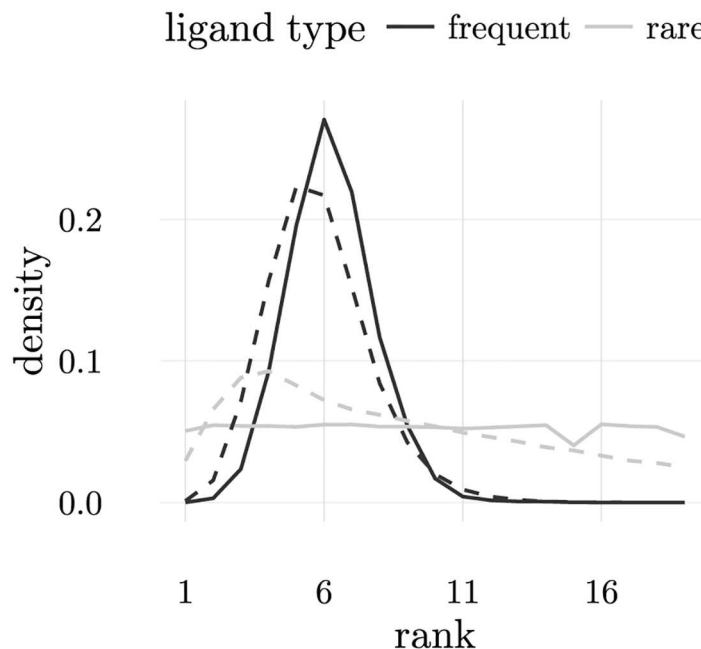


Fig 11. Distribution of the position occupied by frequent and rare LSCSs (dark and light coloured lines) after repertoire education for configurations with $N_r = 0$ (solid lines) and $N_r = 6$ (dashed lines). It is clear from this figure that frequent LSCSs tend to occupy lower positions than rare ligands. When $N_r = 0$, rare ligands are uniformly distributed in ILists, since they do not appear during repertoire education. When $N_r = 6$ the set of ligands appearing during education is larger making it more difficult to order ILists. As a result, in this case frequent LSCSs tend to occupy higher positions in ILists than frequent LSCSs in the $N_r = 0$ case.

doi:10.1371/journal.pone.0169464.g011

lowers, i.e., more LOCS are ranked in top positions. Secondly, the T cell population is considerably heterogeneous with the number of LOCS in top positions ranging from less than 5 to up to 10. This happens in spite of the fact that all T cells are equivalent in terms of their ILists organization. In this case, however, stochastic fluctuations are enough to break down this idealized equivalence. In practical applications, other sources of variation exist since not all APCs present the same information and also because, when limited connectivity is imposed on T cells ILists, T cells do not interact with all APCs.

Negative selection pushes frequent and rare LOCS at different rates. The adaptive nature of T cell repertoire education can also be appreciated by noticing that frequent and rare LSCS tend to occupy different positions in T cells ILists after repertoire education. In particular, the highest ranked rare LSCS tend to be ranked higher than the highest rank frequent LSCS. This happens because negative selection operates more frequently on frequent LSCSs than on rare LSCSs. The paradigmatic example happens in the extreme case in which rare LSCSs do not appear during education, in which case they are nonself. Then negative selection cannot have an impact on how they are ranked in ILists. This can be confirmed in Fig 11, which shows that nonself ligands are ranked uniformly, while frequent LSCS have higher probability of occupying positions away from the top. Rare ligands appear less frequently than frequent LSCSs and consequently they have higher probability of accumulating in higher positions in ILists than frequent LSCSs.

Another important observation concerns the impact of the number of ligands presented on the ordering of ILists after negative selection. The distributions of rare and frequent LSCSs are

centred on higher positions than the distribution of frequent LSCSs when nonself ligands are displayed. In fact, in the first case, the number of ligands that have to be “educated” is higher than when nonself ligands are presented, in which case there is effectively only half of the number of ligands.

There are two sources of limitations that prevent reaching perfectly ordered ILists. The first is statistical: the probability of randomly drawing a perfectly ordered IList decreases exponentially with the list size. The second originates from the intricate frustrated dynamics. When a T cell establishes a long conjugation with an APC, it prevents other T cells from interacting with the same ligand. As a result some of these T cells can be eliminated by negative selection since they sense a higher number of absent LOCS to this indirect effect. Every time a T cell is eliminated by negative selection, it is replaced by another naïve cell with a random IList, and consequently, in the whole population there is a high probability that it may be eliminated again in another upcoming configuration. Furthermore, every time a T cell establishes stable conjugations with some LSCS it may send other T cells to education as a side effect, making it hard to reach a state where all T cells have a similar probability of maintaining tolerance.

Generalisation capabilities depend on the presentation frequency. Another factor contributing to the creation of fluctuations during education is the rate of change of the information presented by APCs (presentation frequency). In our simulations, the configuration of ligands presented by APCs changes every T_S iterations. T_S has an important effect in the ordering of ILists and in detection. For very large values of T_S , education strictly avoids having LSCS on top positions in any IList. For small values of T_S , to be eliminated during education, it is required that in two consecutive configurations a T cell lacks frequent LOCS on top positions in its IList. Therefore, LSCSs, and in particular, rare LSCSs can appear on top positions in some ILists because the probability that they appear in consecutive configurations is p_R^2 which is small ($p_R^2 \sim 2\%$ when $p_R = 15\%$). Consequently, when T_S is small there will be more LOCS in top positions, on average, even if there may be some rare LSCSs among them. This agrees with the sequence of results in Fig 12).

Applications

In the previous sections it was shown that CFSs can perform elaborate discrimination tasks. Now we evaluate how these discrimination capabilities compare to the performance of well-known statistical tests. We end by considering even more complex scenarios for which one must resort to data mining algorithms.

Can the immune system perform a t-test?. Probably the best known and most widely used statistical test is the t-test. For this reason we questioned whether cellular systems could perform this test with comparable accuracy. To define self-configurations, ordered samples with 80 numbers were randomly drawn from a Gaussian distribution with an average of $\mu_S = 50$ and standard deviation $\sigma = 10$. Abnormal self-configurations were obtained by drawing samples from Gaussian distributions with the same σ , but with $\mu_{NS} = 50 \pm \Delta$ (two sided tests). The i^{th} APC displayed the i^{th} number in the sample as a ligand. Therefore, the first APC always presented the smallest number while the last APC, the highest. T cells sense these ligands as R or F ligands, depending on whether the number falls inside or outside an interval where a fraction v_j of the numbers displayed by an APC in normal configurations lie. Here, j is the T cell index, and $v_j < 50\%$ since it corresponds to the frequency of rare ligands.

In this article, T cells define rare ligands intervals for all ligands either on the right or on the left tail. Their size is controlled by the discrimination parameter v_j which is drawn from a uniform distribution between 0 and v_{max} . As a consequence, T cells with $v_j \approx 0$ are only sensitive to nonself ligands.

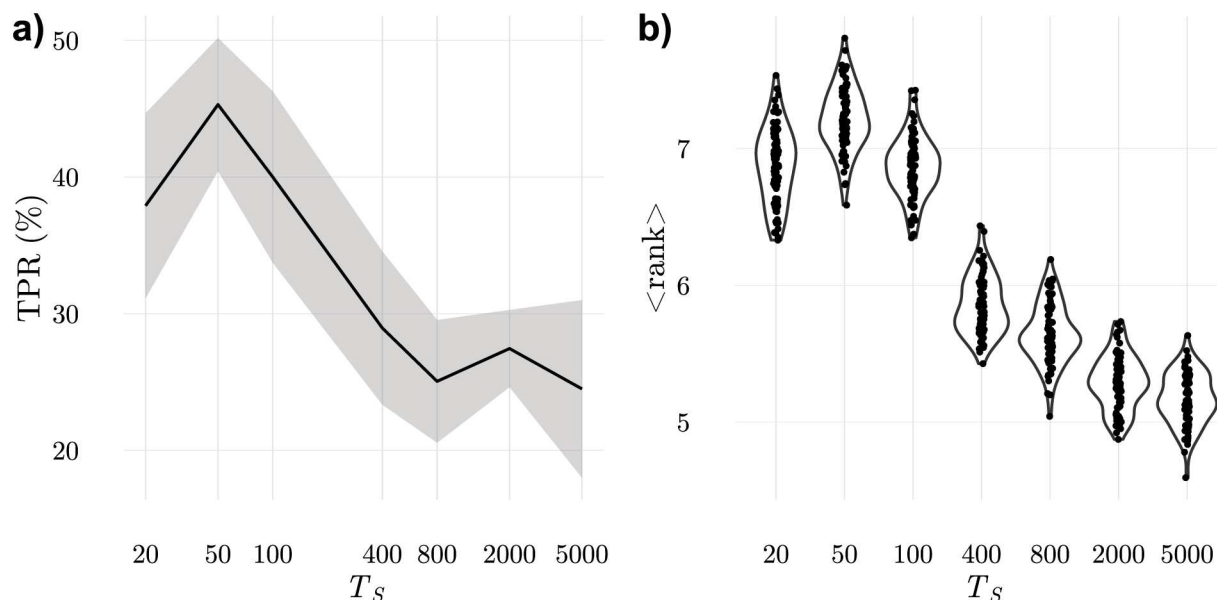


Fig 12. Generalisation capabilities are gained when samples are changed after every short time interval, T_s . a) true positive rate for discrimination of configurations with an added rare ligand ($N_r = 7$) b) mean rank of the highest ranked LSCS in T cell ILists in the 200 samples used for education. These results show that ordering of ILists is best achieved when samples are changed every $T_s = 50$ iterations which also leads to the best discrimination. Since τ_n —the maximum conjugation duration used to eliminate T cells by negative selection—is changed only when no cell is eliminated in the last $W_r = 10000$ iterations, this forces ILists to be consistent with the last $10000/50 = 200$ samples.

doi:10.1371/journal.pone.0169464.g012

To assess how CFSs detect changes in μ_S , the T cell population underwent negative selection, during which 1000 self-configurations were presented by APCs. Afterwards, APCs presented another 1000 self-configurations and 1000 abnormal self-configurations. In Fig 13, receiver operating characteristic curves (ROC curves) are presented for different changes in μ , with $\Delta = \pm 1, \pm 2, \pm 3$. Comparison with the results that would be obtained with a t-test or a KS-test are also provided.

The best results are obtained with a t-test, as expected because data was prepared respecting its assumptions, and also because the t-test is a uniformly most powerful test for detecting deviations on the distribution mean [56–59]. By contrast and as expectable, the KS-test has the lowest performance since it is a simple nonparametric test. Results obtained with CFSs are, to a certain extent surprising since little assumptions were necessary on the type of information displayed by APCs and on the type of changes that could take place.

The CFF took three important assumptions in this example. The first was that the several elements in a sample could be ordered since all derived from a same Gaussian distribution. The same assumption is used in the KS test. The second assumption consisted in placing rare ligands in the tails. This guarantees that if the distribution deviates to either side the number of rare ligands present in the sample increases, and therefore this change has a high probability of being detected. Finally, the third assumption consisted in assuming that rare ligands in all APCs tend to appear on the same side of the distribution. This assumption makes sense because, given the ordering applied to the elements in the sample, a deviation in one tends to produce similar deviations in the next.

The three assumptions taken are important to obtain the best results, and they show if the immune system uses a cellular frustration strategy to perform discrimination, then evolution

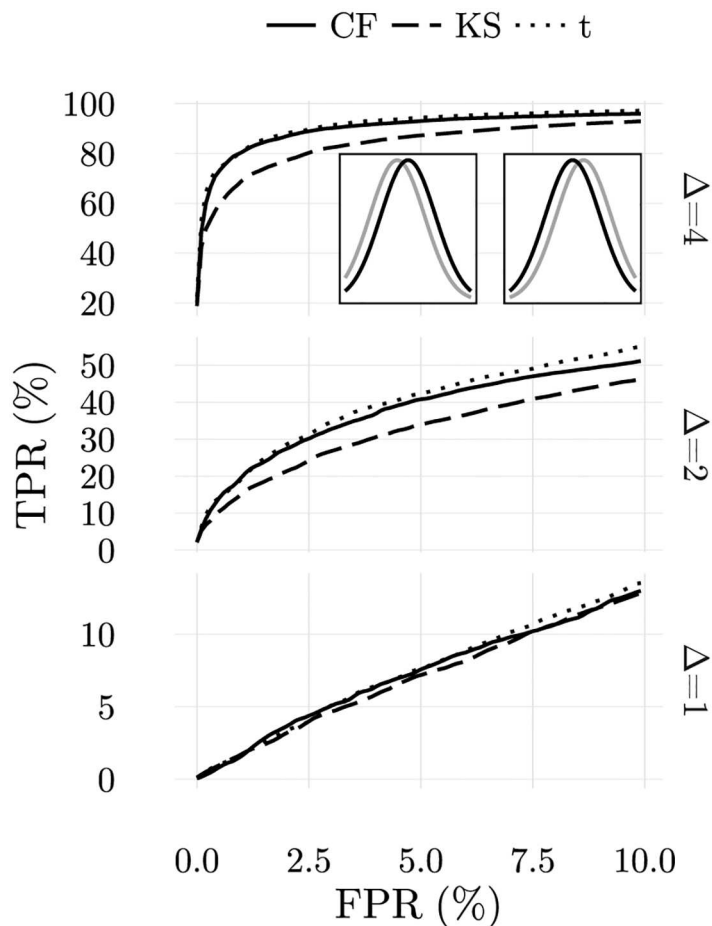


Fig 13. Comparative analysis of average ROC curves obtained in location tests using ordered samples with 80 elements drawn from normal distributions for the cellular frustration model and two-sided KS-test and t-test. Insets: comparison between gaussian distributions used to draw self-configurations (black) and abnormal configurations (black) and abnormal configurations with deviations to either side (gray). Similar plots would be obtained for the distributions in the bottom examples. Abnormal-distributions are only slightly displaced. The t-test is the best estimator as demonstrated in the literature for this ideal case. However, cellular frustration models give very close results.

doi:10.1371/journal.pone.0169464.g013

could have played a role to incorporate similar restrictions that allow reducing the space of potential disturbances. In the present problem, the first assumption makes sense given the nature of information presented. The second and third assumptions, could be relaxed if it would be important to contemplate a wider range of challenges. The same type of trade-offs distinguish parametric and nonparametric tests in statistics. In this sense, CFSs work as non-parametric tests, and are more accurate than KS tests, although computationally more involved. In fact, an interesting link exists between the KS approach and the CFF. In the KS test, the maximal deviation from a standard cumulative distribution determines whether the null-hypothesis should be abandoned. This decision crucially depends on sampling ordering. In the CFF the same deviations appear in the form of missed frequent ligands. The advantage of CFSs is to account to all deviations and furthermore to the presence of nonself ligands.

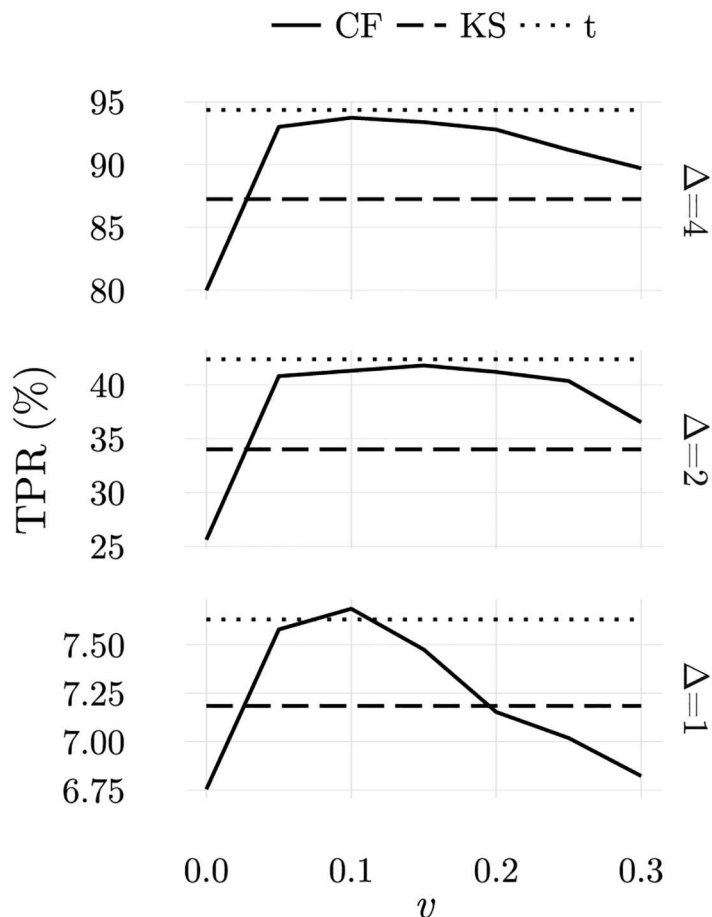


Fig 14. Average true positive rates obtained when the maximal probability of perceiving the information displayed by an APC as a rare signal (or ligand) is v_{max} for the three location tests conducted. Discrimination is best for $v_{max} > 0$ which means that discrimination is not exclusively due to the presence of nonself ligands. A 5% false positive rate was considered.

doi:10.1371/journal.pone.0169464.g014

Another issue that should be addressed consists in knowing whether CFSs are robust if non-optimal parameters, such as v_{max} , are used. This is particularly important because the optimal parameters typically depend on the challenge. In Fig 14 we show how results for a false positive rate of 5% change when v_{max} is changed. Clearly, when v_{max} is in the range $[0.05, 0.15]$, results are consistently good for the different perturbations. This result is interesting because it shows that the best results are obtained when a fraction of cells perform abnormal-self detection, and not simply the detection of nonself ligands, i.e., outliers. From these results, we can conclude that the CFF contains the necessary mechanisms to tackle discrimination tasks conceptually equivalent to those solved by t-tests, an interesting result given the differences in the methods.

Can the immune system perform robust statistics? An important area of research in statistics is concerned with improving the performance of statistical methods when the conditions required for their application are not met. CFSs are computationally more involved, so one could question if they offer robust solutions. We use the simple and paradigmatic example of the performance of t-tests in location tests when samples derive from a log-normal

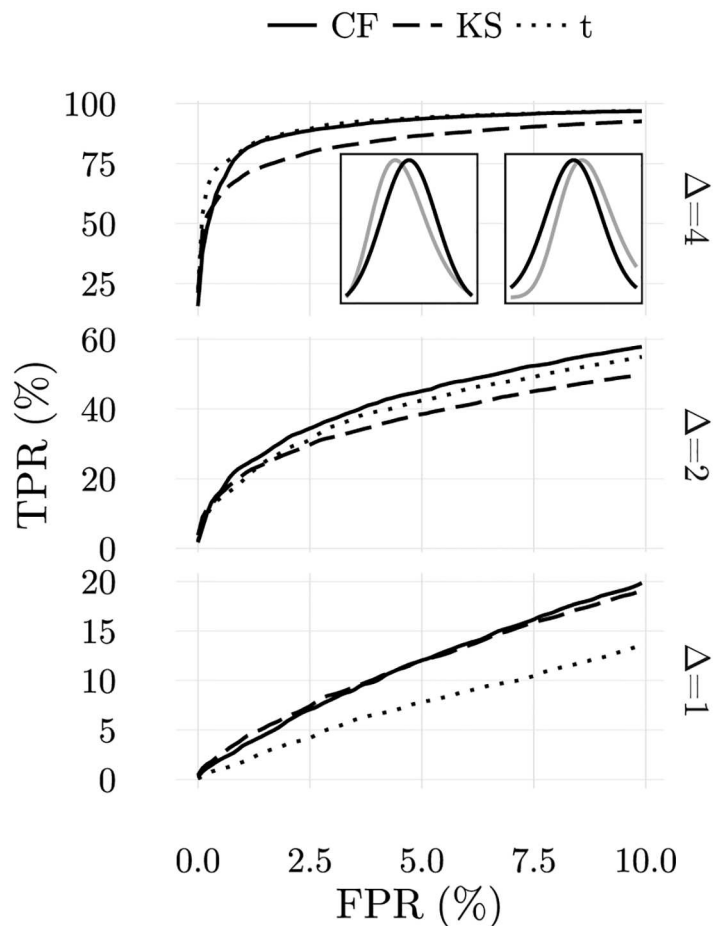


Fig 15. Comparative analysis of average ROC curves obtained in location tests using ordered samples with 80 elements drawn from lognormal distributions for the cellular frustration model and two-sided KS-test and t-test. Insets: comparison between the lognormal distributions used to draw self-configurations (black) and abnormal configurations with deviations to either side (gray). Cellular frustration models outperform the other statistical tests.

doi:10.1371/journal.pone.0169464.g015

distribution. This case is challenging to the t-test because, even though, according to the central limit theorem, the distribution of the mean should converge to the normal distribution, the presence of heavy tails makes this convergence slow. This is particularly true for small samples. Consequently, the t-test can fail considerably.

To test the accuracy of CFSs in this case, 1000 ordered samples with 80 numbers were drawn from log normal distributions with the same means and standard deviations as in the previous example. Furthermore, $v_{max} = 0.05$ was used and the same mapping into frequent and rare ligands was chosen. The ROC curves obtained are shown in Fig 15. CFSs demonstrate robustness over the range of perturbations considered.

Is the immune system a sophisticated data miner? In more realistic scenarios, many pieces of information have to be accounted before triggering an immune response. Each of them can depend on different sources of variation and consequently a multivariate analysis is required. A simple way to mimic this more general scenario is to assume that in the previous examples samples are not ordered, i.e. each element in a sample is drawn from independent

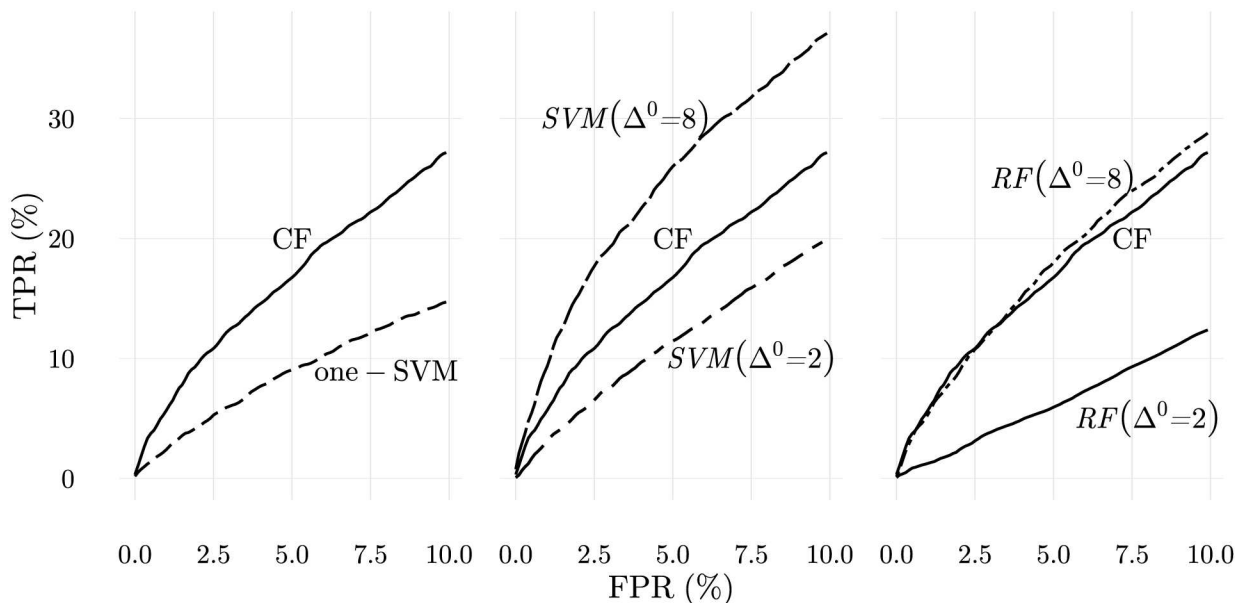


Fig 16. Comparison of the performance of the cellular frustration (CF) model with state of the art data mining algorithms. On the left the results obtained from CF model are compared with one-class SVM. The CF model clearly outperforms the data mining algorithm. In order to confirm that the results from the CF model are realistic, the same results are compared with support vector machines (center) and random forests (right) classification algorithms. Classification algorithms need to be trained with samples from abnormal configurations (drawn from Gaussians with mean deviated by Δ^0), and consequently the comparison with the CF model, which uses only information from self-configurations, would be unfair. However, it shows that even with extra information, classification algorithms can fail to produce good results. This happens when the examples of abnormal configurations are not sufficiently distinct from those belonging to self-configurations. See text and Methods for more details.

doi:10.1371/journal.pone.0169464.g016

generators. Consider next the example in which, in self-configurations, each APC presents a ligand drawn from a Gaussian distribution centred at $\mu_S = 50$ and with standard deviation $\sigma = 10$. Furthermore, in abnormal configurations μ_S is changed to $\mu_{NS} = 50 \pm 4$. These changes occur independently for each APC and in each configuration.

To evaluate CFSs responses, the same strategy used to map APCs ligand information into rare or frequent ligands and to select the T cell repertoire is taken, and ROC curves calculated. To evaluate the performance of CFSs, a comparison with the best data mining techniques for this type of tasks was established (see [Methods](#) for more implementation details of these methods). The results are presented in [Fig 16](#).

One-class SVMs, just like the CFSs defined here, only use self-configurations to infer what is normal and abnormal. Therefore, CFSs and one-class SVMs are methods that use the same information to accomplish the same task. The results shown in [Fig 16](#) show that CFSs perform considerably well, outperforming one-class SVMs. Note nevertheless, that our goal here is not to argue that this may always be the case. In fact, SVMs solve an optimization problem and consequently datasets could be designed for which they provide the best solutions. Our goal is to demonstrate that it is not hard to find relevant problems for which CFSs are better, at least when standard approaches are used (i.e., using standard sets of parameters – see [Methods](#)).

Since one-class SVMs performances were so poor in comparison to CFSs, another set of numerical simulations was conducted to compare CFSs with classification methods, like two class SVMs and random forests (RFs). Classification methods require both, normal and abnormal configurations for training. Therefore, a direct comparison with CFSs is not straightforward. However, classification methods can indicate whether the information captured by CFSs

and depicted in Fig 16 is indeed present, and how classification methods behave when the information provided for training does not perfectly match the information required for detection.

In Fig 16 a comparison of the performance of two class SVMs and random forests (RFs) with CFSs is presented. In these results, classification methods were trained with two sets of samples of the same size and generated from Gaussian distributions centred at $\mu = 50 \pm \Delta^0$. For self-configurations, $\Delta^0 = 0$, while for abnormal self-configurations $\Delta^0 = 2$ or $\Delta^0 = 8$. For detection, configurations centred around $\mu = 50 \pm \Delta$, with $\Delta = 0$ and $\Delta = 4$, were used. These results show that if classification methods have access to abnormal configurations that do not strikingly contrast with self-configurations ($\Delta^0 = 2$), then discrimination of abnormal configurations centred around more distant positions ($\Delta^0 = 4$) can be problematic. In brief, classification methods work well only if distinctive examples are consistently presented. This, of course, can be problematic if only mildly abnormal self-configurations are available. In more mundane terms one could argue that classification methods are only good at predicting the occurrence of earthquakes if a strong earthquake of the same type has been felt before.

Finally, these results also show that the immune system can still improve its performance if it learns from examples of abnormal self-configurations. This, we know, the immune does, through clonal expansion mechanisms, a topic that we leave for a future publication.

Conclusions and Perspectives

The results reported in this paper represent a landmark in the cellular frustration framework and its relevance to immunology because they clearly show that through simple cellular interactions the immune system could be activated with high precision.

The CFF proposed that cellular frustrated interactions could be crucial to build a system of cells in interaction whose dynamics would be extremely sensible to changes in population configurations [42]. This happened because generalized kinetic proofreading mechanisms were identified [42, 43, 60] making each T cell a context discriminator (or detector). Indeed, in [60] it was shown that in CFSs the dissociation constant that distinguishes alternative pathways before major signals are produced [61] is renormalized by factors related to cell frequencies in the population. The production of major signals was straightforwardly related to long-lived conjugations and it was proposed that the focus for understanding how the adaptive immune system works should be placed on which cells establish those long-lived conjugations.

In [49] it was shown that, despite the fact that CFSs are very sensitive to changes, a set of configurations could be defined to be tolerated. To select which configurations to tolerate, a new organization principle was proposed involving all interaction lists (ILists) in the population. The principle of maximal frustration suggested that ILists were selected to reduce conjugation lifetimes for all cells in the population and in this way taking into account the information in normal (self) configurations.

The previous results motivated a research program that had as main goal identifying the simplest frustrated system that could explain fundamental observations and simultaneously why the immune system is competent at protecting the host. In [48] it was shown that the simplest system could involve only APCs and T cells. This would be enough to explain why the immune system accurately detects nonself ligands in the periphery. However, to accomplish this, negative and positive selection would be required. Furthermore, costimulation and anergy would be necessary to improve the accuracy. These results are consistently accounted within the CFF scenario, as summarized in Fig 2.

By raising the question “Can the immune system perform a t-test?” this work calls attention to the fact that performing self-nonself discrimination could be largely insufficient to detect

anomalies in the immune system. In statistics it is well known that the null hypothesis can be rejected even in the absence of outliers. It can be the presence of multiple small biases that signal an anomaly. Here we showed that CFSs could perform the same tasks as today's most popular statistical and machine learning techniques, and with comparable, if not greater, accuracies. This can have profound immunological consequences since it was shown how the immune system can monitor simultaneously the presence of nonself ligands and the presence of combinations of frequently presented ligands.

The current results are part of a framework, a different mind setting to model cellular interactions and detect changes in the displayed information. These results suggest new directions of research, many of them naturally linked to immunology. Indeed, B cell activation was not considered, although the activation of B cells by T helper cells in lymph nodes [62, 63] may likely be explained using a similar modelling approach. Investigating how B cells are activated can also be extremely relevant because the considerable data that exists today [64, 65] makes it possible to confront theory and observation [66–68]. In this respect, it would be important to include the impact of effector functions like clonal expansion and the introduction of regulatory T cells to study how the immune response regulates homeostasis. To understand the immune system the CFF proposes finding the simplest model that is capable of performing the most accurate anomaly detection under plausible immunological conditions. For instance, in [49] we showed that three cell types (T cells, APCs and Tregs) could build a frustrated dynamics with detection capabilities. However, in [48] we showed that to achieve perfect self-nonself discrimination, Tregs were not required. Likewise, here we show that a model with only T cells and APCs is enough to explain how abnormal detection capabilities can be achieved. Therefore, so far, the CFF suggests that Tregs play no essential role for achieving good detection capabilities. This conclusion applies at least under static conditions. However, this conclusion may change if other requirements are imposed on how the information displayed evolves in time. This will be a matter of discussion in a future publication.

Another important conclusion resulting from this work concerns the role played by frequent and rare ligands for activation of the immune system. In [48] it was shown that nonself ligands—the rarest ligands possible—trigger specific cell activations, when these ligands appear in the system. By contrast, in abnormal-self discrimination, long-lived conjugations do not necessarily involve rare ligands and arise due to the absence of sets of frequent ligands. Therefore, the immune system monitors the two types of ligands in the system and in different ways.

Another point worth noting is that the results reported in this paper show how the immune system can be triggered to respond to changes in the displayed information. However, nothing is said concerning how responses can be set to resolve the causes of challenges. This is another topic to be discussed elsewhere and it will require an enlarged systemic approach to model the immune system.

How the real information displayed by APCs shapes the information contained in T cell ILists (i.e., T cell receptors) is a topic that should be further investigated taking into account the real aminoacid frequencies [53], details on recombinations events [69] or details on the interaction energies between APCs and T cell receptors [70].

Another issue that could be interesting exploring is the relation between CFF to other frustration models used in immunology [71–74] and possible connections to neuronal networks [75]. In particular, it would be interesting how the two classes of models compare at performing equivalent immune protection tasks.

Finally, the ideas explored in this article can be further developed in other directions. It would be interesting to consider whether the concept-drift idea [76] proposed to address anomaly detection in non-stationary datasets in artificially immune systems could also be

useful, to explain the adaptive immune system. Another possible direction, would be to understand how danger theory ideas [24, 77] could be compatible with the CFFs. This work also suggests several ideas for research in machine learning, such as extending the current results to unsupervised learning [78] and even deep learning strategies in anomaly detection [79].

Methods

Agents Based Model

The agent based model used in this article can be summarized as follows. The model considers two cell types, APCs and T cells, each having two subtypes, I and II , with $N/2$ cells each. APCs favour interactions with T cells of the same subtype. Each APC presents a single ligand, p_i . For an APC with index i , p_i is a real number between i and $i + 1$. Each T cell has a connectivity list with K different APCs with whom it can interact with. T cells map the information sensed on each APCs into only two possible signals (or ligands), F_i or R_i , standing respectively for frequent or rare. Different specific mapping rules can be designed. In this article each T cell either senses rare ligands on the right or the left tail. For each T cell with index j a discrimination parameter v_j is drawn from a uniform distribution between 0 and v_{max} . Designate $f_i^0(x)$ the frequency of occurrence of ligand p_x in self-configurations presented during education. Define left and right discrimination ligand thresholds, p_i^L and p_i^R , such that:

$$\int_i^{p_i^L} f_i^0(x) dx = v_j, \quad \int_{p_i^R}^{i+1} f_i^0(x) dx = v_j \quad (8)$$

Then, a ligand p_i is mapped onto a rare ligand R_i if $p_i < p_i^L$ for T cells sensing rare ligands on the left tail, or $p_i > p_i^R$ if they sense them on the right tail. In the DinBs case study, all T cells used the same mapping and therefore each APC can be associated to a rare or frequent ligand. In this case, p_i 's were selected so that the number of rare ligands in each configuration was fixed. In more general cases, p_i derive from the dataset.

Each T cell arranges F_i and R_i ligands in ordered interaction lists (ILists), prioritizing interactions with top ranked ligands (or signals). Modelling of cellular interactions assumes a discrete time dynamics where, at each time step, a randomly drawn cell is put in interaction with a cell from the other cell type and belonging to its connectivity list. A new conjugation is established whenever the two cells that are put in interaction, prioritize this interaction. In that case, if they were already conjugated, former conjugations are terminated and the duration of that conjugation is registered (see S1 Fig).

Negative selection in thymic repertoire education is modelled eliminating T cells that remain conjugate for a time longer than a threshold lifetime τ_n (see S1 Fig). A new cell is introduced in the population with the same connectivity list but with a randomly drawn IList. If after W_τ iterations (typically 10000 iterations) no cells exceeded the threshold time, then τ_n is updated to the largest conjugation time in the last W_τ iterations and the T cell population is registered. Every T_s iterations (typically 100 iterations), ligands presented by APCs present information from a different sample (configuration). When τ_n stops changing for at least 10^6 iterations, education stops, the last recorded population is added to the T cell repertoire and education of a new T cell population is started, maintaining each T cell connectivity list but randomly drawing its IList.

After creating a repertoire of educated T cells, the population enters the calibration stage (see S1 Fig). This should correspond to the final stage of T cell repertoire education. At this stage cells engage in the same decision dynamics as before, except that anergy is introduced, so that T cells conjugated for a time longer than τ_A terminate their conjugations and are replaced

by other cells with the same connectivity in the repertoire. In our results we used $\tau_A = 5$. During calibration only self-information is presented. This is modelled by gathering information for presentation by APCs from the same set of samples that were used in the education stage. During calibration the decision dynamics is run for W_c iterations for each sample (typically $W_c = 10^4$ iterations) and the number of conjugations lasting longer than an activation lifetime τ_{act} involving APC with index i when sample s is presented, $c_{i,s}^0(\tau_{act})$, are registered. Defining the ordered vector $c_{i,(j)}^0(\tau_{act})$, such that $c_{i,(j)}^0(\tau_{act}) \geq c_{i,(j+1)}^0(\tau_{act}) \forall j$, then an activation threshold is established by $n_i^0(\tau_{act}) = c_{i,(x)}^0(\tau_{act})$ where $x = N_s^c \times f$, with N_s^c the number of samples used during the calibration and f is a real number between 0 and 1. Typically we used $f = 0.1$, and hence the 10% biggest number of conjugations lasting a time larger than τ_{act} in a sample was considered. The activation reference time is chosen to be equal to the largest conjugation time in the calibration, i.e., $\tau_{act} = \tau_A$.

To model cellular activation in lymph nodes the decision dynamics is run in the same conditions as in the calibration stage. To analyse the performance of the model in detecting anomalies, APCs display either information from a self-dataset, or from a nonself or abnormal-self dataset. Several examples are illustrated in the Results section. The cellular response to the information displayed by sample s is calculated according to:

$$R_s = \sum_i (c_{i,s}(\tau_{act}) - n_i^0(\tau_{act})) / c_{i,s}(0) \times \theta(c_{i,s}(\tau_{act}) - n_i^0(\tau_{act})) \quad (9)$$

where θ is the Heaviside function. Thus the cellular response sums the increments on the number of long conjugations relatively to the calibration stage. In this expression a subtle contribution from positive selection was taken. Instead of simulating the process explicitly, as in [48], we normalized the number of conjugations in the time interval W_c therefore accounting for the impact of positive selection as depicted in Fig 2.

To quantify the detection accuracy we compute the true positive rate for a fixed false positive rate, FPR . To achieve this we create an ordered vector of population responses to the $N_s^{d,s}$ samples in the detection stage displaying self-information, $R_{(i)}^s$, such that $R_{(i)}^s \geq R_{(i+1)}^s \forall i$ and find R_x^s , where $x = N_s^{d,s} \times FPR$. Then the true positive rate becomes $TPR = \#\{R_s^{ns} : R_s^{ns} > R_x^s\} / N_s^{d,ns}$, where R_s^{ns} are the population responses to samples displaying nonself or abnormal self information and $N_s^{d,ns}$ is the number of nonself or abnormal-self samples presented during the detection stage. The true positive rate is thus equal to the fraction of samples displaying nonself or abnormal self information leading to cellular responses greater than R_x^s .

Applications - dataset generation

Three synthetic datasets were generated to evaluate the performance of cellular frustrated systems (CFSs). All datasets are comprised of samples with 80 elements ($N = 80$).

The first two datasets were used for comparison with conventional statistical tests (t-student and KS-test). In the first case, self-samples were drawn from normal distributions with $\mu_s = 50$ and $\sigma = 10$. Abnormal self-configurations were generated from normal distributions with the same standard deviation but with $\mu_{NS} = 50 \pm \Delta$, where $\Delta = 1, 2$ or 4 depending on the example. In the second case, samples were drawn from lognormal instead of normal distributions. Means and standard deviations were changed so that lognormal distributions maintained the same means and standard deviation.

The third data set was obtained as in the first dataset, however no ordering was applied. Furthermore, each element in abnormal-self samples could be drawn from a normal distribution deviated to either side. The example considered in this case used $\Delta = 4$.

In each comparison, sets of samples were comprised of $N_s^c = 1000$ self samples for training, $N_s^{d,s} = 1000$ self samples for evaluation of the false positive rate, and $N_s^{d,ns} = 1000$ to evaluate the true positive rate. When comparisons with data mining classification algorithms were considered, a supplementary set with 1000 abnormal-self samples with $\mu_{NS} = 50 + \Delta^0$ were used for training. Two cases were studied: $\Delta^0 = 2$ or $\Delta^0 = 8$ (Fig 16).

Applications - Statistical and data mining methods

Conventional two sided t and KS tests used the R implementation with known σ . Packages available in R, randomForests [80] and e1071 (libSVM) [81, 82] were used to apply two data-mining algorithms, respectively random forests and support vector machines. Two versions of support vector machines were considered, one-class SVM with a polynomial Kernel, and two-class SVMs with a RBF (radial basis function) Kernel. Both methods were applied using default packages parameters. This is particularly acceptable in case of anomaly detection, since knowledge of the abnormal class is not available. To compute average performances all tests were repeated ten times.

Supporting Information

S1 Fig. Numerical algorithm flowchart.

(PDF)

S2 Fig. Partially ordered interaction lists organization.

(PDF)

S3 Fig. Companion figure for the discussion of the decrease in the discrimination when a population is educated with N_r^0 rare ligands and $N_r^0 + 1$ rare ligands are presented for discrimination.

(PDF)

S4 Fig. Graph showing the evolution of the number of frequent LOCS in top positions in T cell ILIs in the DinBs case study.

(PDF)

Author Contributions

Conceptualization: FVA.

Data curation: BFF.

Formal analysis: BFF FVA.

Funding acquisition: FVA.

Investigation: BFF PM FVA.

Methodology: BFF PM FVA.

Project administration: FVA.

Resources: FVA.

Software: BFF PM FVA.

Supervision: FVA.

Validation: BFF PM FVA.

Visualization: BFF PM FVA.

Writing – original draft: BFF FVA.

Writing – review & editing: BFF FVA.

References

1. Jerne NK. Towards a network theory of the immune system. *Annales d'immunologie*. 1974; 125C(1-2):373–389. PMID: [4142565](#)
2. Cohen IR. *Tending Adam's Garden: Evolving the Cognitive Immune Self*. Waltham, MA: Academic Press; 2000.
3. Burnet FM, Fenner F. *The Production of Antibodies*. Melbourne: Macmillan; 1949.
4. Cohn M, Langman RE. To be or Not to be Riddled? – That is the Question Addressed by the Associative Antigen Recognition Model*†. *Scandinavian Journal of Immunology*. 2002; 55(4):318–323. doi: [10.1046/j.1365-3083.2002.01059.x](#) PMID: [11967111](#)
5. De Boer RJ, Perelson AS. How diverse should the immune system be? *Proceedings of the Royal Society of London B: Biological Sciences*. 1993; 252(1335):171–175. doi: [10.1098/rspb.1993.0062](#) PMID: [8394577](#)
6. Forrest S, Perelson AS, Allen L, Cherukuri R. Self-Nonself Discrimination in a Computer. In: *Proceedings of the 1994 IEEE Symposium on Security and Privacy*. SP'94. Washington, DC, USA: IEEE Computer Society; 1994. p. 202–212.
7. Celada F, Seiden PE. A computer model of cellular interactions in the immune system. *Immunology Today*. 1992; 13(2):56–62. doi: [10.1016/0167-5699\(92\)90135-T](#) PMID: [1575893](#)
8. Stibor T, Timmis J, Eckert C. On the Use of Hyperspheres in Artificial Immune Systems as Antibody Recognition Regions. In: *Proceedings of the 5th International Conference on Artificial Immune Systems (ICARIS-2006)*. vol. 4163 of *Lecture Notes in Computer Science*. Oeiras, Portugal: Springer-Verlag; 2006. p. 215–228.
9. Ji Z, Dasgupta D. Revisiting Negative Selection Algorithms. *Evolutionary Computation*. 2007; 15(2):223–251. doi: [10.1162/evco.2007.15.2.223](#) PMID: [17535140](#)
10. Tauber AI. Immunology and the enigma of selfhood. In: Wise MN, editor. *Growing Explanations: Historical Perspectives on Recent Science*. Durham NC: Durham: Duke University Press; 2004. p. 199–221.
11. Coutinho A, Forni L, Holmberg D, Ivars F, Vaz N. From an Antigen-Centered, Clonal Perspective of Immune Responses to an Organism-Centered, Network Perspective of Autonomous Activity in a Self-Referential Immune System. *Immunol Rev*. 1984; 79(1):151–168. doi: [10.1111/j.1600-065X.1984.tb00492.x](#) PMID: [6235170](#)
12. Coutinho A, Kazatchkine MD, Avrameas S. Natural autoantibodies. *Current Opinion in Immunology*. 1995; 7(6):812–818. doi: [10.1016/0952-7915\(95\)80053-0](#) PMID: [8679125](#)
13. Eichmann K. *The network collective: Rise and fall of a scientific paradigm*. Birkhauser Verlag AG; 2008.
14. Kieber-Emmons T, Monzavi-Karbassi B, Pashov A, Saha S, Murali R, Kohler H. The Promise of the Anti-Idiotypic Concept. *Frontiers in Oncology*. 2012; 2(196). doi: [10.3389/fonc.2012.00196](#) PMID: [23267437](#)
15. Varela FJ, Coutinho A. Second generation immune networks. *Immunology Today*. 1991; 12(5):159–166. doi: [10.1016/S0167-5699\(05\)80046-5](#) PMID: [1878127](#)
16. Schulz R, Werner B, Behn U. Self tolerance in a minimal model of the idiotypic network. *Frontiers in Immunology*. 2014; 5(86). doi: [10.3389/fimmu.2014.00086](#) PMID: [24653720](#)
17. Agliari E, Barra A, Ferraro GD, Guerra F, Tantari D. Anergy in self-directed B lymphocytes: A statistical mechanics perspective. *Journal of Theoretical Biology*. 2015; 375:21–31. doi: [10.1016/j.jtbi.2014.05.006](#) PMID: [24831414](#)
18. Langman RE, Cohn M. A minimal model for the self-nonself discrimination: a return to the basics. *Seminars in Immunology*. 2000; 12(3):189–195. doi: [10.1006/smim.2000.0231](#) PMID: [10910739](#)
19. Tauber AI. The immune self: theory or metaphor? *Immunology Today*. 1994; 15(3):134–136. doi: [10.1016/0167-5699\(94\)90157-0](#) PMID: [8172646](#)
20. Efroni S, Cohen IR. Simplicity belies a complex system: a response to the minimal model of immunity of Langman and Cohn. *Cellular Immunology*. 2002; 216(1-2):23–30. doi: [10.1016/S0008-8749\(02\)00504-X](#) PMID: [12381347](#)

21. Dembic Z. On Recognizing “Shades-of-Gray” (Self-Nonself Discrimination) or “Colour” (Integrity Model) by The Immune System. *Scand J Immunol*. 2013; 78(4):325–338. doi: [10.1111/sji.12090](https://doi.org/10.1111/sji.12090) PMID: [23819602](https://pubmed.ncbi.nlm.nih.gov/23819602/)
22. Burnet SFM. The clonal selection theory of acquired immunity. vol. 3. Vanderbilt University Press Nashville; 1959.
23. Kelsey J, Timmis J. In: Cantú-Paz E, Foster JA, Deb K, Davis LD, Roy R, O'Reilly UM, et al., editors. *Immune Inspired Somatic Contiguous Hypermutation for Function Optimisation*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. p. 207–218.
24. Gu F, Greensmith J, Aickelin U. Theoretical formulation and analysis of the deterministic dendritic cell algorithm. *Biosystems*. 2013; 111(2):127–135. doi: [10.1016/j.biosystems.2013.01.001](https://doi.org/10.1016/j.biosystems.2013.01.001) PMID: [23337179](https://pubmed.ncbi.nlm.nih.gov/23337179/)
25. Aydin I, Karakose M, Akin E. Chaotic-based hybrid negative selection algorithm and its applications in fault and anomaly detection. *Expert Systems with Applications*. 2010; 37(7):5285–5294. doi: [10.1016/j.eswa.2010.01.011](https://doi.org/10.1016/j.eswa.2010.01.011)
26. Dasgupta D, KrishnaKumar K, Wong D, Berry M. Negative Selection Algorithm for Aircraft Fault Detection. In: *Lecture Notes in Computer Science*. Springer Science + Business Media; 2004. p. 1–13.
27. Dasgupta D, Gonzalez F. An immunity-based technique to characterize intrusions in computer networks. *IEEE Transactions on Evolutionary Computation*. 2002; 6(3):281–291. doi: [10.1109/TEVC.2002.1011541](https://doi.org/10.1109/TEVC.2002.1011541)
28. Golovko V, Bezobrazov S, Kachurka P, Vaitsekhovich L. In: Koronacki J, Raś ZW, Wierchoń ST, Kacprzyk J, editors. *Neural Network and Artificial Immune Systems for Malware and Network Intrusion Detection*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 485–513.
29. Idris I, Selamat A. Improved email spam detection model with negative selection algorithm and particle swarm optimization. *Applied Soft Computing*. 2014; 22:11–27. doi: [10.1016/j.asoc.2014.05.002](https://doi.org/10.1016/j.asoc.2014.05.002)
30. Pavone M, Narzisi G, Nicosia G. Clonal selection: an immunological algorithm for global optimization over continuous spaces. *Journal of Global Optimization*. 2011; 53(4):769–808. doi: [10.1007/s10898-011-9736-8](https://doi.org/10.1007/s10898-011-9736-8)
31. Campelo F, Guimaraes FG, Igarashi H, Ramirez JA. A clonal selection algorithm for optimization in electromagnetics. *IEEE Transactions on Magnetics*. 2005; 41(5):1736–1739. doi: [10.1109/TMAG.2005.846043](https://doi.org/10.1109/TMAG.2005.846043)
32. Babayigit B, Akdagli A, Guney K. A Clonal Selection Algorithm for null Synthesizing of Linear Antenna Arrays by Amplitude Control. *Journal of Electromagnetic Waves and Applications*. 2006; 20(8):1007–1020. doi: [10.1163/156939306776930222](https://doi.org/10.1163/156939306776930222)
33. Whitbrook AM, Aickelin U, Garibaldi JM. Idiotypic Immune Networks in Mobile-Robot Control. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 2007; 37(6):1581–1598. doi: [10.1109/TSMCB.2007.907334](https://doi.org/10.1109/TSMCB.2007.907334)
34. Wang YN, Hsu HH, Lin CC. In: Wang L, Chen K, Ong YS, editors. *Artificial Immune Algorithm Based Obstacle Avoiding Path Planning of Mobile Robots*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005. p. 859–862.
35. Stibor T, Mohr P, Timmis J, Eckert C. Is Negative Selection Appropriate for Anomaly Detection? In: *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation. GECCO'05*. New York, NY, USA: ACM; 2005. p. 321–328.
36. Gett AV, Sallusto F, Lanzavecchia A, Geginat J. T cell fitness determined by signal strength. *Nature Immunology*. 2003; 4(4):355–360. doi: [10.1038/ni908](https://doi.org/10.1038/ni908) PMID: [12640450](https://pubmed.ncbi.nlm.nih.gov/12640450/)
37. Bousso P. T-cell activation by dendritic cells in the lymph node: lessons from the movies. *Nature Reviews Immunology*. 2008; 8(9):675–684. doi: [10.1038/nri2379](https://doi.org/10.1038/nri2379) PMID: [19172690](https://pubmed.ncbi.nlm.nih.gov/19172690/)
38. Abbas AK, Lichtman AH. *Basic Immunology: Functions and Disorders of the Immune System*. Philadelphia, PA: Elsevier/Saunders; 2010.
39. Germain RN. The Art of the Probable: System Control in the Adaptive Immune System. *Science*. 2001; 293(5528):240–245. doi: [10.1126/science.1062946](https://doi.org/10.1126/science.1062946) PMID: [11452112](https://pubmed.ncbi.nlm.nih.gov/11452112/)
40. Depoel D, Zaru R, Guiraud M, Chauveau A, Harriague J, Bismuth G, et al. Immunological Synapses Are Versatile Structures Enabling Selective T Cell Polarization. *Immunity*. 2005; 22(2):185–194. doi: [10.1016/j.immuni.2004.12.010](https://doi.org/10.1016/j.immuni.2004.12.010) PMID: [15723807](https://pubmed.ncbi.nlm.nih.gov/15723807/)
41. Beltman JB, Marée AFM, Lynch JN, Miller MJ, de Boer RJ. Lymph node topology dictates T cell migration behavior. *The Journal of Experimental Medicine*. 2007; 204(4):771–780. doi: [10.1084/jem.20061278](https://doi.org/10.1084/jem.20061278) PMID: [17389236](https://pubmed.ncbi.nlm.nih.gov/17389236/)
42. de Abreu FV, Nolte-Hoen ENM, Almeida CR, Davis DM. Cellular Frustration: A New Conceptual Framework for Understanding Cell-Mediated Immune Responses. In: *Lecture Notes in Computer Science*. Springer Science + Business Media; 2006. p. 37–51.

43. Almeida CR, de Abreu FV. Dynamical instabilities lead to sympatric speciation. *Evolutionary Ecology Research*. 2003; 5(5):739–757.
44. Gale D, Shapley LS. College Admissions and the Stability of Marriage. *The American Mathematical Monthly*. 1962; 69(1):9–15. doi: [10.2307/2312726](https://doi.org/10.2307/2312726)
45. Irving RW. The cycle roommates problem: a hard case of kidney exchange. *Information Processing Letters*. 2007; 103(1):1–4. doi: [10.1016/j.ipl.2007.02.003](https://doi.org/10.1016/j.ipl.2007.02.003)
46. Gu Y, Zhang Y, Pan M, Han Z. Matching and Cheating in Device to Device Communications Underlying Cellular Networks. *IEEE Journal on Selected Areas in Communications*. 2015; 33(10):2156–2166. doi: [10.1109/JSAC.2015.2435361](https://doi.org/10.1109/JSAC.2015.2435361)
47. Gusfield D, Irving RW. *The Stable Marriage Problem: Structure and Algorithms*. Cambridge, MA, USA: MIT Press; 1989.
48. Mostardinha P, de Abreu FV. Positive and negative selection, self-nonself discrimination and the roles of costimulation and anergy. *Scientific Reports*. 2012; 2. doi: [10.1038/srep00769](https://doi.org/10.1038/srep00769) PMID: [23101027](https://pubmed.ncbi.nlm.nih.gov/23101027/)
49. de Abreu FV, Mostardinha P. Maximal frustration as an immunological principle. *Journal of The Royal Society Interface*. 2009; 6(32):321–334. doi: [10.1098/rsif.2008.0280](https://doi.org/10.1098/rsif.2008.0280) PMID: [18715844](https://pubmed.ncbi.nlm.nih.gov/18715844/)
50. Parisi G. Two signals from B cells control the expansion of T cells: only one is immunologically specific. *Annales de l'Institut Pasteur / Immunologie*. 1988; 139(2):177–185. doi: [10.1016/0769-2625\(88\)90039-6](https://doi.org/10.1016/0769-2625(88)90039-6) PMID: [3259425](https://pubmed.ncbi.nlm.nih.gov/3259425/)
51. Curtsinger JM, Mescher MF. Inflammatory cytokines as a third signal for T cell activation. *Current Opinion in Immunology*. 2010; 22(3):333–340. doi: [10.1016/j.coi.2010.02.013](https://doi.org/10.1016/j.coi.2010.02.013) PMID: [20363604](https://pubmed.ncbi.nlm.nih.gov/20363604/)
52. Altan-Bonnet G, Germain RN. Modeling T Cell Antigen Discrimination Based on Feedback Control of Digital ERK Responses. *PLoS Biology*. 2005; 3(11):e356. doi: [10.1371/journal.pbio.0030356](https://doi.org/10.1371/journal.pbio.0030356) PMID: [16231973](https://pubmed.ncbi.nlm.nih.gov/16231973/)
53. Košmrlj A, Chakraborty AK, Kardar M, Shakhnovich EI. Thymic Selection of T-Cell Receptors as an Extreme Value Problem. *Physical Review Letters*. 2009; 103(6). PMID: [19792616](https://pubmed.ncbi.nlm.nih.gov/19792616/)
54. Chakraborty AK, Košmrlj A. Statistical Mechanical Concepts in Immunology. *Annual Review of Physical Chemistry*. 2010; 61(1):283–303. doi: [10.1146/annurev.physchem.59.032607.093537](https://doi.org/10.1146/annurev.physchem.59.032607.093537) PMID: [20367082](https://pubmed.ncbi.nlm.nih.gov/20367082/)
55. Chao DL, Davenport MP, Forrest S, Perelson AS. A stochastic model of cytotoxic T cell responses. *Journal of Theoretical Biology*. 2004; 228(2):227–240. doi: [10.1016/j.jtbi.2003.12.011](https://doi.org/10.1016/j.jtbi.2003.12.011) PMID: [15094017](https://pubmed.ncbi.nlm.nih.gov/15094017/)
56. Neyman J, Pearson E. Sufficient statistics and uniformly most powerful tests of statistical hypotheses. 1 SRM, editor. *Statist*; 1936.
57. Neyman J, Pearson E. Contributions to the Theory of Testing Statistical Hypotheses. 2 SRM, editor. *Statist*; 1938.
58. King ML. The power of Students's t test: can a non-similar test do better? *Australian Journal of Statistics*. 1990; 32(1):21–27. doi: [10.1111/j.1467-842X.1990.tb00996.x](https://doi.org/10.1111/j.1467-842X.1990.tb00996.x)
59. Wilcox RR. *Fundamentals of Modern Statistical Methods*. Springer Science + Business Media; 2010.
60. Lindo AM, Faria BF, de Abreu FV. Tunable kinetic proofreading in a model with molecular frustration. *Theory in Biosciences*. 2012; 131(2):77–84. doi: [10.1007/s12064-011-0134-z](https://doi.org/10.1007/s12064-011-0134-z) PMID: [21948153](https://pubmed.ncbi.nlm.nih.gov/21948153/)
61. McKeithan TW. Kinetic proofreading in T-cell receptor signal transduction. *Proceedings of the National Academy of Sciences*. 1995; 92(11):5042–5046. doi: [10.1073/pnas.92.11.5042](https://doi.org/10.1073/pnas.92.11.5042) PMID: [7761445](https://pubmed.ncbi.nlm.nih.gov/7761445/)
62. Goodnow CC, Sprent J, de St Groth BF, Vinuesa CG. Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature*. 2005; 435(7042):590–597. doi: [10.1038/nature03724](https://doi.org/10.1038/nature03724) PMID: [15931211](https://pubmed.ncbi.nlm.nih.gov/15931211/)
63. Goodnow CC, Vinuesa CG, Randall KL, Mackay F, Brink R. Control systems and decision making for antibody production. *Nature Immunology*. 2010; 11(8):681–688. doi: [10.1038/ni.1900](https://doi.org/10.1038/ni.1900) PMID: [20644574](https://pubmed.ncbi.nlm.nih.gov/20644574/)
64. Weinstein JA, Jiang N, White RA, Fisher DS, Quake SR. High-Throughput Sequencing of the Zebrafish Antibody Repertoire. *Science*. 2009; 324(5928):807–810. doi: [10.1126/science.1170020](https://doi.org/10.1126/science.1170020) PMID: [19423829](https://pubmed.ncbi.nlm.nih.gov/19423829/)
65. Cohen IR. Autoantibody repertoires, natural biomarkers, and system controllers. *Trends in Immunology*. 2013; 34(12):620–625. doi: [10.1016/j.it.2013.05.003](https://doi.org/10.1016/j.it.2013.05.003) PMID: [23768955](https://pubmed.ncbi.nlm.nih.gov/23768955/)
66. Madi A, Bransburg-Zabary S, Kenett DY, Ben-Jacob E, Cohen IR. The Natural Autoantibody Repertoire in Newborns and Adults. 2012; p. 198–212.
67. Bransburg-Zabary S, Kenett DY, Dar G, Madi A, Merbl Y, Quintana FJ, et al. Individual and meta-immune networks. *Physical Biology*. 2013; 10(2):025003. doi: [10.1088/1478-3975/10/2/025003](https://doi.org/10.1088/1478-3975/10/2/025003) PMID: [23492831](https://pubmed.ncbi.nlm.nih.gov/23492831/)
68. Mora T, Walczak AM, Bialek W, Callan CG. Maximum entropy models for antibody diversity. *Proceedings of the National Academy of Sciences*. 2010; 107(12):5405–5410. doi: [10.1073/pnas.1001705107](https://doi.org/10.1073/pnas.1001705107)

69. Murugan A, Mora T, Walczak AM, Callan CG. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proceedings of the National Academy of Sciences*. 2012; 109(40):16161–16166. doi: [10.1073/pnas.1212755109](https://doi.org/10.1073/pnas.1212755109)
70. Burroughs NJ, de Boer RJ, Keşmir C. Discriminating self from nonself with short peptides from large proteomes. *Immunogenetics*. 2004; 56(5):311–320. doi: [10.1007/s00251-004-0691-0](https://doi.org/10.1007/s00251-004-0691-0) PMID: [15322777](https://pubmed.ncbi.nlm.nih.gov/15322777/)
71. Parisi G. A simple model for the immune network. *Proceedings of the National Academy of Sciences*. 1990; 87(1):429–433. doi: [10.1073/pnas.87.1.429](https://doi.org/10.1073/pnas.87.1.429) PMID: [2296597](https://pubmed.ncbi.nlm.nih.gov/2296597/)
72. Agliari E, Barra A, Guerra F, Moauro F. A thermodynamic perspective of immune capabilities. *Journal of Theoretical Biology*. 2011; 287:48–63. doi: [10.1016/j.jtbi.2011.07.027](https://doi.org/10.1016/j.jtbi.2011.07.027) PMID: [21824481](https://pubmed.ncbi.nlm.nih.gov/21824481/)
73. Agliari E, Annibale A, Barra A, Coolen ACC, Tantari D. Immune networks: multitasking capabilities near saturation. *Journal of Physics A: Mathematical and Theoretical*. 2013; 46(41):415003. doi: [10.1088/1751-8113/46/41/415003](https://doi.org/10.1088/1751-8113/46/41/415003)
74. Agliari E, Annibale A, Barra A, Coolen ACC, Tantari D. Immune networks: multi-tasking capabilities at medium load. *Journal of Physics A: Mathematical and Theoretical*. 2013; 46(33):335101. doi: [10.1088/1751-8113/46/33/335101](https://doi.org/10.1088/1751-8113/46/33/335101)
75. Agliari E, Barra A, Galluzzi A, Guerra F, Moauro F. Multitasking Associative Networks. *Physical Review Letters*. 2012; 109(26). doi: [10.1103/PhysRevLett.109.268101](https://doi.org/10.1103/PhysRevLett.109.268101) PMID: [23368622](https://pubmed.ncbi.nlm.nih.gov/23368622/)
76. Abi-Haidar A, Rocha LM. In: Bentley PJ, Lee D, Jung S, editors. *Adaptive Spam Detection Inspired by a Cross-Regulation Model of Immune Dynamics: A Study of Concept Drift*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 36–47.
77. Matzinger P. The Danger Model: A Renewed Sense of Self. *Science*. 2002; 296(5566):301–305. doi: [10.1126/science.1071059](https://doi.org/10.1126/science.1071059) PMID: [11951032](https://pubmed.ncbi.nlm.nih.gov/11951032/)
78. Goldstein M, Uchida S. A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. *PLoS ONE*. 2016; 11(4):1–31. doi: [10.1371/journal.pone.0152173](https://doi.org/10.1371/journal.pone.0152173) PMID: [27093601](https://pubmed.ncbi.nlm.nih.gov/27093601/)
79. Kang MJ, Kang JW. Intrusion Detection System Using Deep Neural Network for In-Vehicle Network Security. *PLoS ONE*. 2016; 11(6):1–17. doi: [10.1371/journal.pone.0155781](https://doi.org/10.1371/journal.pone.0155781)
80. Liaw A, Wiener M. Classification and Regression by randomForest. *R News*. 2002; 2(3):18–22.
81. Meyer D, Wien TU. Support Vector Machines: The Interface to libsvm in package e1071. *R News*. 2015; p. 1–3.
82. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. 2011; 2: 2:27:1–27:27. doi: [10.1145/1961189.1961199](https://doi.org/10.1145/1961189.1961199)

Copyright of PLoS ONE is the property of Public Library of Science and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.